DOCUMENT RESUME

ED 456 246 CE 082 222

AUTHOR	Bloom, Howard S.; Hill, Carolyn J.; Riccio, James
TITLE	Modeling the Performance of Welfare-to-Work Programs: The Effects of Program Management and Services, Economic Environment, and Client Characteristics. MDRC Working Papers on Research Methodology.
INSTITUTION	Chicago Univ., IL.; Manpower Demonstration Research Corp., New York, NY.
SPONS AGENCY	Pew Charitable Trusts, Philadelphia, PA.; Department of Health and Human Services, Washington, DC.; Ford Foundation, New York, NY.; Department of Education, Washington, DC.; George Gund Foundation, Cleveland, OH.; Open Society Inst., New York, NY.; New York Times Foundation, NY.; Alcoa Foundation, Pittsburgh, PA.; Ewing Marion Kauffman Foundation, Kansas City, MO.; Ambrose Monell Foundation, New York, NY.; Grable Foundation, Pittsburgh, PA.; California State Dept. of Social Services, Sacramento.; California State Dept. of Education, Sacramento.; California State Job Training Coordinating Council, Sacramento.; Florida State Dept. of Health and Rehabilitative Services, Tallahassee.
PUB DATE	2001-06-00
NOTE	143p.; Also supported by the Anheuser-Busch, Heinz Family, and Union Carbide Foundations.
CONTRACT	97000617-000; HHS-100-89-0030
AVAILABLE FROM	Manpower Demonstration Research Corporation, 16 East 34 Street, New York, NY 10016. Tel: 212-532-3200; Web site: http://www.mdrc.org/. For full text: http://www.mdrc.org/Reports2001/EffectsofPrgmMgmt-WkgPpr/EffectsPrgMgmt-Method.pdf.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Adult Education; *Client Characteristics (Human Services); *Economic Climate; Economic Impact; Economically Disadvantaged; Employment Patterns; *Employment Programs; Predictor Variables; *Program Administration; *Program Effectiveness; Program Evaluation; Program Implementation; Unemployment; Wages; *Welfare Recipients; Welfare Services
IDENTIFIERS	Impact Studies; Randomization (Statistics); *Welfare to Work Programs

ABSTRACT

This paper poses a question of direct relevance for welfare administrators, program operators, and policy makers: What management practices, program strategies, and local conditions are key to running effective welfare-to-work programs? To address this question, the present analysis links detailed measures of program characteristics to valid and precise estimates of program effects on short-term earnings. Data for the analysis are drawn from these three random assignment studies of welfare-to-work programs in 59 sites across the United States (with a combined total of 69,399 welfare clients): California's Greater Avenues for Independence (GAIN) program, Florida's Project Independence (PI), and the National Evaluation of Welfare-to-Work Strategies (NEWWS). The findings indicate that, other things being equal, program effects on earnings during

Reproductions supplied by EDRS are the best that can be made
from the original document.

the first two years after random assignment are largest when programs strongly emphasize employment, provide personalized attention, and do not let staff caseloads become large. The paper also finds that short-term impacts on earnings are smaller where unemployment is high, management choices and practices matter a lot, increased reliance on basic education reduces short-run effects, economic environment plays an important role, and program effectiveness varies inconsistently with client characteristics. Appendixes include 54 references; data tables and figures; program models for GAIN, PI, and NEWWS; and descriptions of measuring program performance as program effects on client earnings, measuring program characteristics, and testing the sensitivity of findings from the impact model. (Contains 54 references.) (Author/YLB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 456 246

MDRC Working Papers on Research Methodology

Modeling the Performance of Welfare-to-Work Programs: The Effects of Program Management and Services, Economic Environment, and Client Characteristics

Howard S. Bloom
Carolyn J. Hill
James Riccio

BEST COPY AVAILABLE

Manpower Demonstration
Research Corporation

MDRC

June 2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

J.S. Gressm

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

CE 082 222

This working paper is part of a new series of publications by MDRC on alternative methods of evaluating the implementation and impacts of social programs and policies.

The research reported in the present paper was funded through a subcontract with the University of Chicago from grant #97000617-000 awarded by The Pew Charitable Trusts. The authors are at MDRC (Bloom and Riccio) and the University of Chicago (Hill).

The authors are grateful to Laurence E. Lynn, Jr. of the University of Chicago, for valuable counsel and support for the present research agenda. For insights into the programs studied, assistance with the data used, and other advice, they thank their colleagues Thomas Brock, Stephen Freedman, Gayle Hamilton, James Kemple, Charles Michalopoulos, and Electra Small of MDRC, and Alan Orenstein.

MDRC's evaluation of the California Greater Avenues for Independence (GAIN) Program was funded mainly by the California Department of Social Services (CDSS), with additional support from the U.S. Department of Health and Human Services (HHS).

MDRC's evaluation of Florida's Project Independence (PI) Program was funded by Florida's State Department of Health and Rehabilitative Services and with additional support from the Ford Foundation and HHS.

MDRC is conducting the National Evaluation of Welfare-to-Work Strategies with funding from HHS under a competitive award, Contract No. HHS-100-89-0030. HHS is also receiving funding for the evaluation from the U.S. Department of Education. The study of one of the sites in the evaluation, Riverside County (California), is funded in part by a contract with CDSS which, in turn, is receiving additional funds from the California State Job Training Coordinating Council, the California Department of Education, HHS, and the Ford Foundation.

Dissemination of MDRC publications is also supported by the following foundations and individuals who help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: the Ford, Ewing Marion Kauffman, Ambrose Monell, Alcoa, George Gund, Grable, Anheuser-Busch,. New York Times Company, Heinz Family, and Union Carbide Foundations; and the Open Society Institute.

The findings and conclusions presented in this paper do not necessarily represent the official positions or policies of the funders.

For information about MDRC, see our Web site: www.mdrc.org.
MDRC® is a registered trademark of the Manpower Demonstration Research Corporation.

Copyright © 2001 by the Manpower Demonstration Research Corporation. All rights reserved.

Abstract

This paper poses a question of direct relevance for welfare administrators, program operators, and policy makers: What management practices, program strategies, and local conditions are key to running effective welfare-to-work programs? To address this question, the present analysis links detailed measures of program characteristics to valid and precise estimates of program impacts on short-term earnings. The data for the analysis are drawn from three random assignment studies conducted by MDRC of welfare-to-work programs in 59 sites across the U.S. (with a combined total of 69,399 welfare clients): California's Greater Avenues for Independence (GAIN) program, Florida's Project Independence (PI), and the National Evaluation of Welfare-to-Work Strategies (NEWWS). The findings indicate that, other things being equal, program impacts on earnings during the first two years after random assignment are largest when programs strongly emphasize employment, provide personalized attention, and do not let staff caseloads become large. The paper also finds that short-term impacts on earnings are smaller where unemployment is high. Future papers will address corresponding questions about other labor market and welfare outcomes in the short and longer term.

Table of Contents

CHAPTER AND SECTION **PAGE**

1. **Introduction** 1
   - 1.1 Improving the Performance of Human Service Organizations 2
   - 1.2 Measuring Performance 2
     - 1.2.1 Specifying Intended Organizational Outcomes 2
     - 1.2.2 Identifying Organizational Contributions to Intended Outcomes 3
   - 1.3 Modeling Performance 4

2. **Analytic Strategy** 7
   - 2.1 Conceptual Framework 7
     - 2.1.1 Program Management Choices and Practices 7
     - 2.1.2 Program Services 12
     - 2.1.3 Economic Environment 14
     - 2.1.4 Client Characteristics 15
   - 2.2 Statistical Model 18
     - 2.2.1 Objectives of the Model 18
     - 2.2.2 Structure of the Model 18
     - 2.2.3 Specification of the Model 20
     - 2.2.4 Planned Future Extensions of the Model 23

3. **The Settings, Sample, and Measures** 25
   - 3.1 The Settings 25
     - 3.1.1 The Local JOBS Offices 25
     - 3.1.2 The Three Evaluations: GAIN, PI, and NEWWS 26
   - 3.2 The Sample 28
     - 3.2.1 Data Sources and Sample Sizes 28
     - 3.2.2 Analysis Sample 29
   - 3.3 The Program Performance Measure 29
   - 3.4 The Program Management Measures 31
   - 3.5 The Program-Induced Service Differential Measures 33
   - 3.6 The Measure of Economic Environment 35
   - 3.7 The Measures of Client Characteristics 35

4. **The Findings** 37
   - 4.1 The Hypotheses Tested 37
   - 4.2 Management and Performance 38
     - 4.2.1 Client Employment Emphasis: Take a Job Quickly 38
     - 4.2.2 Personalized Client Attention: Getting "Close to the Customer" Can Make a Difference 40
     - 4.2.3 Closeness of Client Monitoring: Information Alone is not Enough 40
     - 4.2.4 Caseload Size: Human Resources Matter 41
     - 4.2.5 Consistency Within the Office: Mixed Results 42

CHAPTER AND SECTION **PAGE**

4.3 Services and Performance: Increased Reliance on Basic Education Reduces Short-Run Effects 43
4.4 Economic Environment and Performance: It's Harder to Increase Earnings When Jobs are Tougher to Find 44
4.5 Client Characteristics and Performance 45

5. Implications and Planned Extensions of the Analysis 48
5.1 Implications of the Present Analysis 48
5.2 Planned Extensions of the Present Analysis 49
5.2.1 Issues to be Addressed 49
5.2.2 How These Issues can be Addressed 51

References 55

APPENDICES

APPENDIX SECTION **PAGE**

A. The Program Models for GAIN, PI, and NEWWS 73

B. Measuring Program Performance as Program Impacts on Client Earnings 78
B.1 Linking Program and Control Group Members to Local Program Offices 78
B.2 Measuring Earnings 78
B.3 Estimating Program Impacts 79
B.3.1 The Starting Point: Unconditional Program Impacts 79
B.3.2 The Next Step: Conditional Program Impacts 80
B.3.3 The "Complete" Impact Model: Including Program Characteristics 81
B.4 Assessing the Statistical Significance of the Impact Variation 82
B.5 "Explaining" the Impact Variation 83

C. Measuring Program Characteristics 86
C.1 Constructing the Program Management Measures 86
C.1.1 The Data Source: Local Staff Surveys 86
C.1.2 The Measures 87
C.1.2.1 Service Technology 87
C.1.2.2 Staff Caseload Size 89
C.1.2.3 Inconsistencies in Views Among Frontline Staff and Supervisors 89
C.2 Constructing the Program Service Measures 90
C.2.1 The Data Source: Client Follow-up Surveys 90
C.2.2 The Program Service Measures 91
C.3 Constructing the Measure of the Local Economic Environment 92
C.3.1 Calculation of Office Unemployment Rates 92

APPENDIX SECTION **PAGE**

- C.3.2 An Alternative Unemployment Measure that was Considered 93
- C.3.3 An Additional Economic Indicator that was Considered 93
- C.4 Assessing the Construct Validity of the Office-Level Measures of Program Characteristics 94
- C.5 Assessing the Statistical Significance of the Variation in Program Characteristics Across Offices 95

D. Testing the Sensitivity of Findings from the Impact Model 110
- D.1 Sensitivity of the Estimated Relationships Between *Program Characteristics* and Program Impacts 110
  - D.1.1 Deleting Riverside GAIN and Portland NEWWS 110
  - D.1.2 Deleting Positive and Negative Outlier Offices based on Impact Estimates 111
  - D.1.3 Deleting Offices with Early Random Assignment 112
  - D.1.4 Deleting Offices Based on High and Low Values of Selected Independent Variables 112
- D.2 Sensitivity of the Estimated Relationships Between *Client Characteristics* and Program Impacts 113
  - D.2.1 Deleting Riverside GAIN and Portland NEWWS 113
  - D.2.2 Deleting Positive and Negative Outlier Offices based on Impact Estimates 114
  - D.2.3 Deleting Offices with Early Random Assignment 114
- D.3 Conclusions 114

TABLES AND FIGURES

TABLE PAGE

1 GAIN, PI, and NEWWS 60

2 Sample Sizes 61

3 Estimated Program Impacts on Mean Total Earnings During the First Two Years After Random Assignment By Local Program Office 62

4 Survey Items for the Management Scales Related to Service Technology 64

5 Summary of Local Program Characteristics 65

6 Summary of Service Receipt Rates and Service Differentials 66

7 Client Characteristics 67

8 Effects of Local Program Characteristics on Program Impacts 68

9 Relationships Between Client Characteristics and Program Impacts 69

10 Program Impacts on Other Outcomes to be Examined in Future Research 70

A1 The NEWWS Program Models 77

B1 Independent Variables in the Unconditional, Conditional, and "Complete" Models of Program Impacts 84

B2 Statistical Significance of the Variation in Program Impacts Across Offices 85

C1 Random Assignment Dates and Samples Sizes for the Analysis Sample 96

C2 Staff Survey Sample Sizes 97

C3 Scales and Survey Items for the Service Technology Measures 99

C4 Item Response Patterns for the Service Technology Scales 100

C5 Reliability Assessments for the Service Technology Scales 101

C6 Values of the Program Management Measures for Each Local Program Office 102

TABLE **PAGE**

C7 Sample Sizes and Values of the Service Differential Measures for Each Local Program Office 104

C8 Descriptive Statistics for the Unemployment Rate at Each Local Program Office 106

C9 Correlations Among Measures of Program Management, Program Services, and the Economic Environment 108

C10 Statistical Significance of the Variation in Program Characteristics Across Offices 109

D1 Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Riverside GAIN and Portland NEWWS 115

D2 Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Offices with the Highest and Lowest Impact Estimates 116

D3 Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Offices with Early Random Assignment 117

D4 Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Offices with Highest and Lowest Values of Selected Independent Variables 118

D5 Sensitivity Tests of the Relationships between *Client Characteristics* and Program Impacts: Deleting Riverside GAIN and Portland NEWWS 119

D6 Sensitivity Tests of the Relationships between *Client Characteristics* and Program Impacts: Deleting Offices with the Highest and Lowest Impact Estimates 120

D7 Sensitivity Tests of the Relationships between *Client Characteristics* and Program Impacts: Deleting Offices with Early Random Assignment 121

FIGURE **PAGE**

1 How Welfare-to-Work Programs Affect Client Earnings 71

2 The Distributions of Unconditional OLS Impact Estimates and Empirical Bayes Impact Estimates 72

A1 The GAIN Program Model 75

A2 The PI Program Model 76

Chapter 1
Introduction

The overriding goal of the present study is to help improve the performance of human service organizations. The specific objectives that flow from this goal are threefold: (1) to provide a conceptual and statistical framework for studying the determinants of the performance of such organizations, (2) to illustrate how this framework can be used to study the performance of a particular type of human service organization—welfare-to-work programs, and (3) to report empirical findings about these programs that are substantively important in their own right. To accomplish these goals, the study uses data from randomized experimental evaluations of three major welfare-to-work programs that operated during the 1990s: California's Greater Avenues for Independence (GAIN) program, Florida's Project Independence (PI), and the National Evaluation of Welfare-to-Work Strategies (NEWWS).

Specifically, the present study estimates the effects of program management, program services, economic environment, and client characteristics on the performance (impacts) of the preceding programs in terms of one key outcome—mean client earnings during the first two years after random assignment.[1] Findings from the study—based on randomized experiments in 59 locations with a total of 69,399 program and control group members—suggest that, holding other factors constant (*ceteris paribus*):

> ***Management choices and practices matter a lot.*** Program impacts increase substantially when a program is managed in such a way as to provide a clearer and more consistent focus on quick employment and/or greater staff emphasis on personalized attention to their clients. Program impacts decrease substantially when staff caseloads are large.
>
> ***Increased reliance on basic education reduces short-run effects.*** Offices that rely more on basic education have lower impacts on client earnings during the first two years after random assignment, other factors equal, compared to offices that rely relatively less on basic education services.
>
> ***Economic environment plays an important role.*** Program-induced earnings gains are larger in areas with lower unemployment rates.
>
> ***Program effectiveness varies inconsistently with client characteristics.*** No clear pattern emerges whereby program impacts are consistently larger or smaller for more disadvantaged or less disadvantaged clients. For example, there are no statistically significant relationships between program impacts and clients' age, race/ethnicity, or pre-program earnings. And while impacts are larger than average

---

[1] Future analyses will explore corresponding findings for other labor market and welfare outcomes and over a longer period of follow-up.

for clients with at least a high school degree or GED, they are also larger for clients who had been receiving welfare payments consistently for at least one year prior to the program, and for those with three or more children.

To help motivate, present, and explain these findings and how they were derived, the paper proceeds as follows. The current chapter provides an overview of the policy context of the analysis and introduces the issues to be addressed. Chapter 2 outlines the analytic framework used. Chapter 3 describes the settings, the sample, and the measures to which the analytic framework was applied. Chapter 4 reports the empirical findings that were obtained and Chapter 5 concludes the paper with a summary of its main findings and a discussion of next steps for future research.

1.1 Improving the Performance of Human Service Organizations[2]

Human service organizations can be defined in many ways, but in general, they comprise "formal organizations explicitly designed to process and change people" (Hasenfeld and English, 1974). Such organizations include, among others: schools, employment agencies, welfare agencies, correctional institutions, hospitals, and mental health clinics. Although the place of human services in modern society is fairly secure, public complaints about their inadequate performance and calls for their reform are as old as the institutions themselves.

Improving the performance of human service organizations is a difficult task, in large part because rigorous empirical evidence is sorely lacking on the question of "what works best for whom?" If social scientists are to help fill this information gap, they must make considerable progress on at least two challenging fronts: (1) improving the measurement of performance, and (2) increasing knowledge about what drives performance. Clearly, it is not possible to know whether performance has improved if one cannot measure it. Equally clearly, it is very difficult to improve performance if one does not know what affects it.

1.2 Measuring Performance

Measuring the performance of a human service organization requires a clear specification of the organization's intended outcomes and an ability to identify its contribution to those outcomes.

1.2.1 Specifying Intended Organizational Outcomes

The first requirement—the existence of clearly specified outcome measures—is much easier to meet for a profit-making firm than it is for a human service organization, which typically is a government agency or a non-profit entity. Among for-profit firms, the primary benchmark of success—financial profit—is both objective and easy to measure.[3] Among human service organizations, however, success is much harder to gauge.

---

[2] The remainder of Chapter 1 draws heavily on an earlier paper on this project (see Riccio, Bloom, and Hill, 2000).

Part of this difficulty stems from the multiple goals that exist for human service organizations. For example, schools are expected to develop general literacy, numeracy and reasoning skills, impart knowledge of subject areas, and help socialize young people. Similarly, correctional institutions often are expected both to separate dangerous individuals from the rest of society (incarcerate them) and to reduce the likelihood of their committing future crimes (rehabilitate them). Likewise, welfare-to-work programs often strive to increase the employment of participants, reduce their dependence on welfare, reduce their poverty, and improve the quality of their lives and the lives of their children.

In practice, these multiple goals often conflict with each other (for example, reducing welfare receipt can increase poverty if not accompanied by corresponding earnings gains) and different stakeholders (program participants, service providers, and taxpayers) often do not agree on the relative priority of goals. Thus, it is difficult, if not impossible, to capture the performance of a human service organization with a universally acceptable "bottom-line" measure.

Measuring performance is also difficult when the desired outcomes are partly or wholly intangible. For example, standardized tests of student achievement, rates of recidivism among former prison inmates, and quality-of-life questions on surveys of former welfare recipients often fail to reflect fully the complex nature of the educational, rehabilitative, and life improvement goals they are intended to represent.

1.2.2 Identifying Organizational Contributions to Intended Outcomes

Even for tangible outcomes that can be measured precisely and objectively, data limited to an organization's own clients are not enough to gauge its performance, because such data cannot isolate the unique contribution that the organization makes to those outcomes. This contribution—often referred to as the *impact of* or *value added by* an organization—is the *change* in outcomes caused by the organization. Measuring this impact requires not only data on client outcomes but also corresponding information about what those outcomes would have been *without* help from the organization. This latter condition is often called a "counterfactual." Ideally then, the performance of an organization should be measured as the *difference* between its actual client outcomes and the counterfactual.

Such evidence is very difficult to obtain, especially in real time on an ongoing basis for an operating program. Although the program evaluation literature is replete with methods for assessing program effectiveness (Cook and Campbell, 1979), it is widely acknowledged that the most valid approach for doing so is a randomized experiment (Betsey, Hollister, and Papageorgiou, 1985). Using this approach, individuals targeted for services are randomly assigned (through a lottery-like process) to either a program group that is offered assistance or a control group that is not. Because the assignment process is random and all sample members have the same chance of being selected for the program, systematic differences in the measured and unmeasured characteristics of the two groups are eliminated, if the sample is sufficiently large.

---

[3] Even for-profit firms have other important goals, such as increasing their stock price or market share. Thus, they have more than one measure of success.

Because the subsequent outcomes of the control group accurately reflect what the program group's outcomes would have been without the program, the difference between the two groups' outcomes provides a valid estimate of the organization's effect, or impact. For example, if 50 percent of a program group in a welfare-to-work initiative found employment compared with only 40 percent of the control group, the 10 percentage point difference in their outcomes represents the impact or added value attributable to the program.

Because of their ability to provide valid program impact estimates, randomized experiments are now widely used to evaluate many types of human service organizations and have played an especially prominent role in the evaluation of employment and training programs (Friedlander, Greenberg, and Robins, 1997; Greenberg and Shroder, 1997).

1.3 Modeling Performance

To date, randomized experiments have been more successful at *documenting* the effectiveness of human service programs (or lack thereof) than they have been at *explaining* why these programs are or are not effective. However, if future performance of programs is to be improved, a better understanding is needed of what accounts for success. This will require going beyond the design of previous experiments, which almost exclusively test the effects of whole programs, not their constituent parts.

This limitation is often referred to as the "black box" problem. Unpacking the black box requires determining how impacts are affected by the nature of the services provided, the manner in which they are implemented, the types of clients who receive them, and the environment in which they are provided.[4]

Multi-site experiments that measure program impacts for each site and collect data on site-related factors that are hypothesized to affect these impacts offer perhaps the best opportunity for studying the determinants of program performance (Greenberg, Meyer, and Wiseman, 1994). With this information it is possible to model the variation in impacts across sites in terms of corresponding differences in site and sample characteristics.[5] The feasibility of such analyses has increased considerably in recent years given advances in the statistical theory of hierarchical models (Bryk and Raudenbush, 1992) and the development of software to estimate these models (Raudenbush et al., 2000).[6]

---

[4] Sherwood and Doolittle (1999) discuss the use of implementation research to explain the results of impact analyses.

[5] Riccio and Orenstein (1996) and Riccio and Hasenfeld (1996) use this approach to study the GAIN program in California.

Although promising, this approach is not foolproof. For one thing, even though evidence of each site's performance (impact) is based on a randomized experiment, statistical models used to estimate relationships between impacts and site and sample characteristics are non-experimental. Hence, they are subject to the same uncertainties that are inherent in any non-experimental analysis.

In particular they are vulnerable to *specification error*, which can occur when important variables are left out of a model.[7] When this happens, part of the influence on program impacts of the left-out variables is attributed to the variables that are included in the model. This will bias estimates of the effect of the included variables. The direction of such biases can be positive or negative, and hence, they can cause model-based estimates to overstate or understate the influence of any given factor on program impacts.

Furthermore, given the complexity and cost of conducting a randomized experiment in many different locations, most past experiments have been confined to a small number of sites. This seriously limits the number of factors whose independent relationships to program impacts can be examined. In addition, it increases the potential for specification bias due to left-out variables, because degrees of freedom must be conserved in empirical models.

The present paper uses a comparison-of-sites strategy to develop a hierarchical model of the relationships between impacts of welfare-to-work programs and four types of program features. Two strategies are used to address (but not eliminate) the inherent limitations of such models. First, to deal with the problem of specification error due to left-out variables, the present study draws on an unusually rich database with information on a variety of program dimensions. Including multiple program measures in the model reduces the number of potentially important left-out program variables. Second, to deal with the small-number-of-sites problem, the present analysis uses a pooled sample from three large multi-site experiments conducted by the Manpower Demonstration Research Corporation (MDRC). This combined sample of experiments provides information for a substantial number of sites (59 local program offices) plus data for large client samples at each site (averaging a total of 1,117 program and control group members per site).

The GAIN, PI, and NEWWS experiments predate the 1996 Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) and the Temporary Assistance for

---

[6] As discussed in section 2.2.2, these models are also known as mixed models, random effects, random coefficient, or variance component models. Other software packages used to estimate these models include SAS Proc Mixed, Stata glamm6, and VARCL, among others.

[7] Greene (1997) pp. 399-404 and Pindyck and Rubinfeld (1998) pp. 184-187 discuss this well-known problem in econometrics.

Needy Families (TANF) program that it created (the cornerstone of recent welfare reform).[8] Nevertheless, these experiments provide the best existing way to study the effects of characteristics of mandatory welfare-to-work programs on their performance. Only by pooling data from a series of large-scale, multi-site experiments is it possible to identify the relationships between program characteristics and valid estimates of the impacts they created. And, to our knowledge, only these three experiments provide consistent detailed quantitative information about how the programs being studied were managed and delivered. Thus, unlike virtually all past research on the determinants of social program performance, the current study focuses directly on factors that lead to *program-induced* changes in client outcomes, not just the outcomes themselves. Furthermore, the types of relationships analyzed are not unique to the pre-TANF era, and thus most likely generalize to current welfare programs and realities.

---

[8] The two-year follow-up period for members of the present analysis sample ranges from March 1988 to June 1992 for GAIN, from January 1991 to August 1993 for PI and from June 1991 to December 1996 for NEWWS.

Chapter 2
Analytic Strategy

The analytic strategy for this study involves the use of a somewhat complicated statistical model that is based on a simple conceptual framework.

2.1 Conceptual Framework

Figure 1 illustrates the conceptual framework for the present analysis. It hypothesizes that at least four factors can directly affect the impacts of welfare-to-work programs on clients' labor market and welfare outcomes: (1) how programs are managed (specifically, the choices that managers make about key features of the organization and its intervention strategies), (2) the services or activities in which clients participate, (3) the conditions of the local labor markets within which the programs operate, and (4) the socio-economic characteristics of the program's clients. (As noted below, these factors may also affect impacts indirectly and/or jointly, but such effects are not explored in the current study.)

For decades, policymakers, program administrators, program staff members, and researchers have argued about the relative importance of these factors. However, the empirical evidence that exists on this topic is quite limited because there has been no comprehensive analysis of *how each factor affects program impacts while holding the other factors constant.* The current study provides such an analysis, and is thus an example of research using multiple levels of information that Lynn, Heinrich, and Hill (2001) argue can provide the most useful insights for public sector governance.

2.1.1 Program Management Choices and Practices

Assessing the influence of management practices and institutional structure on program effectiveness is the focus of two recent studies of employment and training programs for low-income persons funded by the Job Training Partnership Act (JPTA). Heinrich and Lynn (2000) examine the role of Private Industry Councils (which control local job-training programs) and the roles of performance incentives that have been developed for these programs. Using a multi-level or hierarchical modeling approach to take advantage of both client-level and office-level information, they find that *centralized decision making structures* and the *use of performance-based contracts* are associated with better client outcomes (measured by earnings and employment rates). Although their analysis links institutional structure and management practices to program *outcomes,* it does not attempt to link them to program *impacts.*

This next step is taken by Heinrich (2001), who uses similar methods and models to estimate the effects of these same management practices and institutional structure variables on JTPA program impacts. While the coefficient estimates are somewhat different in her models for impacts and outcomes, Heinrich concludes that the fundamental importance of structural and management considerations remains apparent in both.

Another way of thinking about the role of management in welfare-to-work programs concerns the potential for administrators or managers to shape how their frontline workers (e.g., case managers) interact with the agency's clients and the social climate and culture of the office within which many of those interactions take place. Several qualitative studies argue that such factors may be among the most powerful determinants of a program's success. For example, Mead has examined the extent to which program staff members expect clients to work and the efforts they put forth to help clients realize that goal. He concludes that "whether work programs can really *expect* work of their clients may be more critical to their success than the material conditions affecting employment" (1983, p. 649), and that "What goes on inside an office seems to matter even more than the conditions around it"(1986, p. 150). Based on his field research in local offices of early welfare-to-work initiatives funded through the federal Work Incentive Program (WIN), Mead notes that staff in effective offices (based on employment outcomes):

> developed "procedures within procedures" in order to move clients into jobs. They "hustled" to "take care of business."...Because they worked hard, clients did too. Recipients were constantly coming and going on job interviews. The atmosphere was upbeat. You could feel the electricity just walking into these offices. The poor offices, on the other hand, were deathly quiet. Staff were more concerned with following procedures, less with placing clients. (Mead 1986, p.152)

Behn (1991) also paints a striking picture of the inner workings of what he believed to be effective welfare-to-work offices operating Massachusetts' former Employment and Training (ET) Choices program. His "description of management innovations …emphasizes the importance of establishing a clear mission focused on jobs, marketing the program to clients, workers, and the public, convincing both workers and clients that they can in fact achieve the mission, providing them the resources to do so, expecting hard work and energy, and monitoring performance" (paraphrase by Bane, 1989, p. 287).

Similarly, based on her assessment of existing research and first-hand experience as a state welfare administrator, Bane concluded that with respect to promoting effective welfare-to-work programs, "The question then, is how to shape an organizational culture that …delivers a clear message that the goal is jobs, sets a clear expectation that clients can get jobs and that workers are obligated to make that happen, monitors performance, and provides necessary resources" (Bane, 1989, p. 287).

Although the descriptions from these studies may be compelling, the researchers did not have direct evidence about the impacts produced by the programs they studied. Hence, they could not make a convincing empirical link between the management practices they observed and program effectiveness.

Subsequent research by Riccio and Hasenfeld (1996) and Riccio and Orenstein (1996) provide a first step in this direction. These authors use results from a randomized experimental study of the GAIN program in California to compare program impacts with implementation factors across counties and local program offices. Although the small number of GAIN sites (six counties and 20 local offices) limited the number of factors that they could consider, the authors provide some evidence suggesting that how organizations are managed—in particular, the kinds of frontline staff behaviors that managers cultivate—makes a difference in program effectiveness.[1]

Building on the findings of this past research, the present study examines the influence of six organizational factors that reflect or relate to the behavior of frontline staff, and that program managers can influence. This focus is in keeping with a perspective of organizational research that recognizes that frontline workers—sometimes referred to as "street-level bureaucrats"—have substantial influence in determining the organization's *actual* policies toward clients. As Lipsky argues, "Street-level bureaucrats make policy in two related respects. They exercise wide discretion in decisions about citizens with whom they interact. Then, when taken in concert, their individual actions add up to agency behavior" (1980, p. 13).

In the world of welfare-to-work programs, policymakers and administrators establish regulations, guidelines, and reward systems that attempt to define and shape programs that, for example, focus on building basic reading and math skills, building specific occupational skills, promoting rapid entry into the labor force, or a mix of these strategies. However, the ways that frontline staff execute their roles and responsibilities will affect whether the program's *stated* policies become its *operative* policies. Consequently, the nature of the "treatment" that clients experience in welfare-to-work programs depends very much on the behavior of frontline staff, not just the activities in which the clients participate. The present study examines the following set of staff behaviors and related organizational conditions, which are thought to influence program effectiveness:

(1) *The degree of a program's emphasis on moving clients into jobs quickly.*

One important decision for program managers is how much to emphasize the goal of moving registrants into the labor market quickly (even if it means taking a low-paying job). Proponents of a "quick job entry" approach argue that almost any job is a positive first step and that advancement will come through acquiring a work history and learning skills on the job. Consequently, they favor assigning clients to job search assistance as an initial program activity. However, efforts to promote quick employment can pervade staff-client interactions in a variety of ways; the approach is about much more than just assigning clients to job search activities. This can be seen, for example, in the efforts of California's Riverside County, whose unusually large program impacts in the GAIN evaluation helped to popularize this approach:

---

[1] In a recent paper, Dehejia (2000) reexamines data from the GAIN evaluation using multilevel or hierarchical modeling methods. He finds that "most of the differences across sites in treatment impacts are accounted for by differences in the composition of participants" (p. 12). However, when attempting to explain the unusually large impacts produced by the Riverside GAIN offices, he suggests that much of their success may be due to qualitative factors related to program implementation. Because his analysis only focuses on client characteristics, he cannot examine the influences of implementation practices. He therefore suggests that a useful extension of his analysis would be to include characteristics of local program administration and local labor markets, which are key features of the present study.

> Most distinctive was Riverside's attempt to communicate a strong "message" to all registrants (even those in education and training activities), at all stages of the program, that employment was central, that it should be sought expeditiously, and that opportunities to obtain low-paying jobs should not be turned down. The county's management underscored this message by establishing job placement standards as one of several criteria for assessing staff performance, while at the same time attempting to secure the participation of all mandatory registrants. In addition, the county instituted a strong job development component to assist recipients in gaining access to job opportunities (Riccio, Friedlander, and Freedman, 1994, p. xxv).

Supporters of an alternative strategy, referred to as the "human capital development" approach, contend that recipients would do better to participate first in education and training in order to improve their skills and secure new credentials. They should not quickly settle for low-paying jobs, which may not lead to better opportunities. Instead, they should raise their human capital so that they can get better jobs in the future.

(2) The degree of personalized attention given to clients.

A second important feature of the local service technology included in the present analysis reflects the extent to which frontline staff members get to know their clients' personal situations, needs, and goals; arrange service assignments that support these individualized needs and goals; continue to work with clients over time to assure their success in these activities; and adjust client plans to accommodate their changing needs.

Emphasis on personalized client attention does not necessarily imply that staff members ignore program rules. However, the two extremes of these choices are described by "people-processing" versus "people-changing" service technologies (Hasenfeld, 1983):

> On the one hand, a people processing technology emphasizes assessing and classifying clients and assigning them to various service categories. Staff activities focus on collecting and processing information about clients and using that information in accordance with prescribed rules.... A people-changing technology, on the other hand, emphasizes attitudinal and behavioral changes. Staff activities center on developing relationships with clients in order to modify their behavior (Hasenfeld and Weaver 1996, p. 240).

Some administrators believe that a more personalized approach is a more effective strategy than one in which clients are handled in a more routine fashion. An analog in the commercial world would be the notion of "getting close to the customer" to understand and respond to their individual aspirations and situations.

(3) The closeness of client monitoring.

Opinions differ about the importance of participation mandates for the success of welfare-to-work programs (for example, see Bane, 1989; Mead, 1989; Riccio and Hasenfeld, 1996). Supporters contend that such mandates can prompt clients to participate in (and thereby benefit from) program services that they might not receive otherwise. Critics argue that mandates create adversarial relationships between staff and clients that may be inimical to the kinds of changes that programs are seeking to promote. But regardless of the desirability of mandates, if a program attempts to implement one, it will be hard-pressed to enforce it without an effective way to *monitor* client participation. Close monitoring can thus be seen as a necessary condition for enforcement.

Close monitoring is not only important to the enforcement process, but also to staff efforts to *help* clients participate fully in the program and get the most out of the services it offers. Through careful monitoring, staff can determine whether clients are regularly attending and progressing in their activities, whether they need help with personal or situational problems that might be interfering with their attendance and progress, and whether an alternative activity assignment or new support services arrangements are needed to help them participate more successfully. (For similar reasons, careful monitoring is important in voluntary programs as well.)

Because close monitoring may contribute to a program's performance in various ways, it is hypothesized that offices that rate higher on the monitoring scale will have larger impacts on clients' earnings.

(4) and *(5) Frontline staff inconsistency and staff/supervisor inconsistency in views about the agency's service approaches.*

The preceding three measures capture potentially important features of the service technology for welfare-to-work programs. In addition, they provide a way to address another issue that is viewed by many to be crucial for managing effective human services—a clear and consistent message about goals and practices.

Organizational performance may suffer when staff members are divided—whether due to confusion or disagreement—over what the organization should be doing and how it should be doing it. Thus, it has been frequently hypothesized that program administrators can play a critical role in focusing staff effort on a common purpose by instilling a "strong culture" (Behn, 1991;

Miller, 1992; Nathan, 1993). According to this hypothesis, the more successful managers are at instilling a commonality of purpose and views, the more successful their organizations will be.

(6) Staff caseload size.

Large caseloads may prevent caseworkers from spending enough time with clients to be effective (Gueron and Pauly, 1991). Indeed, because of prevailing views about the likely importance of this program feature, the Riverside GAIN site took part in a separate randomized experiment to compare impacts of two different caseloads, one averaging 97 clients per worker (the "normal" caseload for that program) and the other averaging 53 clients per worker (a reduced caseload). That experiment showed that Riverside GAIN's already large impacts on earnings and AFDC payments were no greater for the program group assigned to case managers with the smaller caseloads (Riccio, Friedlander, and Freedman, 1994). To date, this finding remains the only direct existing evidence about the effect of caseload size on the impacts of welfare-to-work programs.

2.1.2 Program Services

Welfare-to-work programs typically offer a variety of employment-related activities and support services. These include: (1) individual and group job search assistance (including job clubs) designed to help connect welfare recipients with existing jobs, (2) basic education (including adult basic education, GED preparation, high school degree programs, and classes in English as a Second Language), (3) classroom training or (less frequently) on-the-job training in specific vocational skills, (4) temporary unpaid work experience (or "workfare") positions designed to help clients with little job experience make a transition into the world of work, and (5) subsidies for child care and transportation to facilitate recipients' participation in program activities. Frontline workers help to arrange for these activities and services, refer clients to the program divisions or other institutions that provide those services, help clients navigate through those institutions, and monitor their participation in their assigned activities.

The present study focuses on the influence of the first three of these program elements: job search assistance, basic education, and vocational training. The relative effectiveness of these activities has been the subject of debate for decades and, as previously mentioned, is part of the broader debate between two fundamentally different philosophies about how best to promote economic self-sufficiency among welfare recipients: quick job entry versus human capital development.

It should be noted that part of the appeal of jobs clubs and individual job search assistance is that they are relatively inexpensive. Thus, its proponents argue, such services can be provided to large numbers of persons within limited government budgets. Education and training are more expensive and may take longer to produce results. Advocates for those services argue, however, that their delayed labor market impacts will be larger and more enduring than those produced by an approach emphasizing rapid employment and will eventually offset the higher costs. These costs reflect the resources necessary to provide high-quality education and training plus the opportunity cost to clients of time spent in class (and thus not earning a wage).

Findings from the GAIN evaluation have been one key point of reference in this debate. Although Riverside County provided a mix of job search assistance and basic education and allowed some use of other education and training opportunities, it more strongly emphasized job search as the preferred initial program activity than did the other study counties, even for recipients who could, under GAIN's rules, choose basic education first. Compared to the other counties, Riverside had by far the largest impacts, and these effects were achieved for the broadest range of client subgroups, including those who entered the program without a high school diploma or GED (Riccio, Friedlander, and Freedman, 1994). This advantage was maintained for five years after random assignment (Freedman, Friedlander, Lin, and Schweder, 1996). More recent findings suggest that Riverside's impacts declined substantially by the seventh year after random assignment (although it is uncertain how much this decline is due to the ending, after the fifth year of follow-up, of the embargo that prevented GAIN from serving members of the control group) (Hotz, Imbens, and Klerman, 2000). Nevertheless, the relative cost-effectiveness of the Riverside GAIN program, which was both less expensive than the others and produced much larger impacts for the first five post-enrollment years, was many times that of any of other GAIN site. However, whether the emphasis on upfront job search was indeed a critical determinant of Riverside's higher performance has remained uncertain because a variety of features, not just the activity mix, distinguished that county's program from the programs operated by the other study counties.

To address the limitation of almost all previous attempts to measure the relative effectiveness of different types and mixes of program activities, NEWWS conducted side-by-side experimental comparisons of a labor force attachment program (emphasizing rapid employment) and a human capital development program in three sites: Riverside, California; Atlanta, Georgia; and Grand Rapids, Michigan. This was done by randomly assigning sample members in each of those sites to one of the two programs or to a control group (Freedman, et al., 2000).

For sample members who entered the study *without* a high school diploma or GED, the human capital development strategies emphasized assignment to basic education activities as the initial step in the program. In contrast, the labor force attachment strategy made upfront job search the first activity. When the early results of the two approaches were directly compared, the labor force attachment strategy proved more effective in raising employment and earnings and reducing the receipt of welfare within the first two years after random assignment.

Among sample members who entered already having a high school diploma or GED, the human capital development strategy usually emphasized assignment to vocational training or post-secondary education as the initial activity, while the labor force attachment program made upfront job search the first activity. Within the first two years of follow-up, the impacts of the human capital development approach were no larger—and were sometimes smaller—than the effects of the labor force attachment strategy. However, the cost of the human capital development strategy was considerably higher. Corresponding analyses for a five-year follow-up period, which would allow a better opportunity for assessing whether human capital development strategies are more effective in the longer term, are currently underway.[2]

When studying the effect of program services, it is important to consider the fact that even without a special program, many welfare recipients will seek and receive employment and training services on their own (Gueron and Pauly, 1991). In part, a program may achieve employment and welfare impacts by increasing the degree to which members of the program group receive such services, compared to what they would have received on their own, as measured by the control group's experiences. Conversely, a program may have little effect on labor market outcomes if it leads to little or no difference in service receipt. The present analysis thus examines the relationship between program impacts and program-generated *service differentials* (i.e., the program group-control group differences in rates of participation in selected activities). A separate service differential measure is computed for job search assistance, basic education, and vocational training.

2.1.3 Economic Environment

That the local economic environment can affect the performance of welfare-to-work programs seems almost self-evident. Nevertheless, there are two diametrically opposed views about the expected direction of this effect.

One view is that program performance is likely to be *better where unemployment rates are lower* (i.e., in tighter labor markets) than where unemployment rates are higher (i.e., in looser labor markets). The argument for this is as follows. With low unemployment rates, there are more job openings for welfare recipients to fill. Therefore, if a program can motivate and prepare additional recipients to seek and qualify for employment, a greater proportion of them will find and take jobs than would be the case if unemployment rates were high and there were fewer available job openings.

The opposing view is that program performance is relatively *worse where unemployment rates are lower*. The argument for this derives from the expectation that where unemployment rates are lower and thus the demand for workers is higher, it is easier for welfare recipients to find jobs even without the help of a program; thus, even though the program may have higher placement rates, the program may actually offer its clients little extra advantage in the labor market. This especially may be the case among recipients who are the most job-ready. At the same time, recipients who cannot find jobs where unemployment rates are lower may have personal characteristics or situational barriers that make them harder to employ. If this is the case, it will be harder than otherwise for a program with limited resources per client to increase employment.

---

[2] For findings covering a five-year follow-up period for an earlier generation of mandatory welfare-to-work programs that focused primarily on job search assistance and unpaid work experience, see Friedlander and Burtless, 1995.

A second version of this argument appeals to the intuition of "ceiling effects." It posits that the larger the proportion of a group that finds employment on its own is, the smaller the margin will be for any program to make a difference. This argument is only plausible, however, when the underlying counterfactual is near the relevant ceiling, which often is not the case for welfare-to-work programs—especially those for long-term recipients, whose likelihood of full-time employment can be well below 50 percent and who typically get only low-paying jobs.

The current empirical basis for assessing these competing views is extremely limited given the very small number of prior systematic attempts to compare site-level program impact estimates from randomized experiments to corresponding measures of the local economic environment. Furthermore, the few previous attempts to do so (for example, Riccio and Orenstein, 1996) are based on small numbers of sites, which seriously limits the statistical power of their comparisons and their ability to control for other site-level factors.

To address this issue, the present study develops a measure of the prevailing unemployment rate faced by clients in each of the 59 local program offices and includes this measure in a model of program impacts.

2.1.4 Client Characteristics

There are many plausible reasons to expect that human service programs in general, and welfare-to-work programs in particular, will perform differently for different types of participants. Moreover, there are at least three important reasons for wanting to know how this performance varies: (1) to help target program resources, (2) to help set standards for program performance, and (3) to help interpret the findings of program evaluations.

Decisions about targeting resources are a major concern for all human service programs. Equity and efficiency are two criteria that often conflict for making these decisions. The equity criterion is often stated in terms of providing services to clients in accord with their *need for assistance*. The efficiency criterion is often stated in terms of providing services in accord with clients' *ability to benefit*. To understand the implications of these criteria it is important to know how program performance differs for different types of clients.[3]

---

[3] Smith and Plesca (forthcoming) summarize current attempts to target employment services using statistical "profiling" models that predict the likely need for or benefit from these services. In addition they review the limited and mixed existing empirical evidence on the ability of such models to increase program impacts. The only large-scale profiling system that is operational currently is the Worker Profiling and Reemployment Services program used by many state Unemployment Insurance agencies to assign mandatory services to unemployment insurance benefit claimants who are judged likely to receive benefit payments for a very long time or to exhaust the benefits to which they are entitled.

How program performance varies for different types of clients also has important implications for setting performance standards for managing the delivery of human services. This is especially critical because those services are usually provided through decentralized delivery systems. Thus, in order to ensure that services are provided fairly and effectively, and that the organizations providing them are held accountable, it is necessary to assess the performance of the multiple agencies and agents involved. However, this leads immediately to the following conundrum: how to compare success across agencies when different organizations serve different types of clients, some of whom may be more difficult to serve than others. If this comparison is not done properly, organizations may have powerful incentives to target clients who are easiest to serve. This well-known and widely documented phenomenon is often referred to as *creaming*, and how to address it remains a thorny public management challenge.[4]

In contrast, organizations with a particular ideological commitment may decide to target difficult-to-serve persons who are most in need of help. For example, Heckman, Smith, and Taber (1996) find that JTPA caseworkers in Corpus Christi, Texas showed preferences for serving clients with the greatest needs, contrary to the incentives offered by the performance standards system. Such organizations can be seriously penalized for making this commitment if future funding is based on performance standards that do not properly account for their client mix. To help develop such systems it is essential to understand how client characteristics affect program performance.

A third important reason for needing information about how program performance varies for different types of clients arises in the context of evaluations of programs that produce small or no impacts on average, but substantial impacts for policy-relevant subgroups. One striking example is a recent randomized experimental study of Career Academies, an innovative high school reform program. This study found that Career Academies reduced dropout rates substantially for students who were at high risk of academic failure, but had little or no impact for the much-larger group of other students (Kemple and Snipes, 2000). When all students were averaged together, however, the large effect for high-risk students was obscured. Thus, it was essential to examine how student (client) characteristics were associated with program impacts.

The client characteristics judged to be most relevant for welfare-to-work programs are those representing *barriers to employment*, or conversely, *employability*. The most widely used indicators for this purpose are formal education, prior employment experience, and past welfare receipt. Formal education and prior employment experience represent human capital; past welfare receipt predicts future reliance on welfare. Other indicators that have been used include race/ethnicity (to reflect potential labor market discrimination or other impediments specific to certain groups), number and age of children (to reflect alternative demands on clients' time), physical health status, and mental health status.

---

[4] For decades there has been widespread speculation about the extent to which creaming is stimulated by performance standards based on client outcomes. However, the empirical evidence on this issue is limited and mixed (Heckman, Heinrich, and Smith, 1999).

Past researchers have successfully linked these measures to program outcomes such as future levels of employment, earnings, and welfare receipt. However, they have been far less successful in linking them to measures of program impacts.

The most influential attempt to do so is Friedlander's (1988) study of the relationships among program outcomes, program impacts, and client employability using data from randomized experiments conducted on five welfare-to-work programs from the early 1980s, most of which were relatively low-cost interventions that emphasized job search and work experience activities. Friedlander found below average or no earnings impacts for sample members who were the most job-ready or the least job-ready. In contrast, he found that sample members who were in the middle range of the employability distribution experienced the largest program-induced earnings gains (although these gains were modest).

Friedlander's findings suggest that: (1) the most job-ready clients may have been the best able to find jobs on their own and thus had the smallest margin for improving in response to the limited services being offered, (2) the least job-ready clients may have had the greatest margin for improving but the least ability to change their situation, and (3) the middle group may have had the best balance (for the programs being studied) between its margin for improvement and its ability to improve. As Gueron and Pauly (1991, p. 157) note, the Friedlander study "provided strong evidence against 'creaming'—i.e., serving only the most advantaged, who demonstrate high placement rates—but did not confirm narrow targeting of these low- to moderate-cost programs on the most disadvantaged."[5]

Michalopoulos, Schwartz, and Adams-Ciardullo (2001) paint a different picture from their recent subgroup analysis of later-generation randomized experiments, but one that also argues against narrowly targeting welfare-to-work programs toward any particular subgroups. They find that "Overall, the programs increased earnings and reduced welfare payments for most subgroups" (p. ES-5), and that these programs "increased earnings about as much for the most disadvantaged groups as for the moderately and least disadvantaged groups" (p. ES-10). At the same time, not every program was equally effective for all of the subgroups it served. Indeed, the study showed that programs that followed a "mixed service" strategy that included at least some opportunities for clients to choose between job search and education or training activities as their initial activities were effective for the broadest range of subgroups.

To build on this previous research, the present study examines the relationships between program impacts on future earnings and the following client characteristics: their education level, recent past employment and earnings experience, recent past welfare receipt, age, race/ethnicity, and number and age of children.

---

[5] In contrast to his findings for earnings impacts, Friedlander (1988) found that program-induced welfare reductions were largest for sample members who were the least employable.

2.2 Statistical Model

To help determine how each of the preceding factors influences program impacts, the present study specifies their relationships as a statistical model and estimates the parameters of this model from data for 59 local welfare-to-work program offices. The objectives, structure, specification, limitations, and some likely future extensions of the model are described below.

2.2.1 Objectives of the Model

The primary objective of the model is to provide *estimates of the independent effects of program, environmental, and client characteristics on program impacts on individuals' earnings while holding constant (controlling for) the effects of the other characteristics.* Thus, for example, the model is designed to provide estimates of the effect of a specified office management factor on program impacts while holding constant the influences of other management features, program services, economic environment, and client characteristics.

A second major objective of the model is to account properly for the use of the site (i.e., local program office) as well as the individual welfare recipient as units of analysis. The facts that individual welfare recipients are clustered within sites, and that the analysis seeks to estimate cross-site statistical relationships among site-level variables as well as among individual-level variables, have important implications for tests of the statistical significance of the model's parameter estimates. Thus, the process for assessing statistical significance must take account of these conditions.

A third major objective of the model is to account properly for the difference between how *estimates* of program impacts vary across local program offices and how their underlying *"true"* impacts vary. This distinction acknowledges the fact that the impact for each office is observed (estimated) with error. Hence, there are two sources of between-office variation in observed impacts: (1) estimation error, and (2) differences in true impacts. To properly estimate the relationships between program characteristics and program impacts requires identifying these two variance components.

2.2.2 Structure of the Model

The preceding three objectives can be met by specifying the present analysis as a *hierarchical linear production function.* Production functions are a standard analytic device used by economists to examine relationships between the inputs and outputs of a production process. Recent versions of the approach are referred to as value-added models (Meyer, 1997). Among its many applications, this approach has been used to study the linkages between educational inputs (characteristics of students, teachers, parents, and schools) and educational outputs (measures of student achievement). Barnow (1979) provides a detailed review of the early educational production function literature and extends the approach to employment and training programs. Following a production function approach, the present analysis conceptualizes program management, program services, economic environment, and client characteristics as inputs to a production process with program impacts on future earnings as its output.

Hierarchical models (also called random effects models, mixed models, or variance components models) are a major advance in the analysis of data where observations are clustered within aggregate units. These units themselves might be grouped within higher-order aggregate units. This applies, for example, to: students within classes within schools within school districts; employees within establishments within firms; residents within families within neighborhoods; pre-tests and post-tests for the same individuals; and longitudinal data with multiple observations for each sample member. For the present analysis, the application of interest is program and control group members within local offices of welfare-to-work programs.

Bryk and Raudenbush (1992) provide perhaps the best-known and most complete discussion of hierarchical modeling; Raudenbush et al. (2000) is perhaps the most widely used hierarchical modeling software package. One of the most popular current applications of the approach is the estimation of educational production functions as value-added models (Bryk, et al., 1998; Raudenbush and Willms, 1995; and Sanders and Horn, 1994).

For the present analysis, a two-level hierarchical linear model is used to specify a production function relating program impacts (outputs) to program, environmental, and client characteristics (inputs). The unit of analysis for Level One is the individual sample member; the unit of analysis for Level Two is the local program office.

Level One is a linear regression equation that is based on data for individual sample members and includes a separate conditional impact for each site, controlling for its client characteristics.[6] Estimating this equation serves two purposes. First, it provides estimates of the effects of client characteristics on program impacts, which is of direct substantive interest. Second, the conditional impact estimates for each local office provide the key dependent variable for Level Two of the model.

Level Two contains three regression equations. One equation provides the core substantive findings for the present study; the other two are necessary to complete the model, but are of less substantive interest. The first equation specifies the conditional program impact for each office as a function of its program management, program services, and economic environment. Thus, its parameter estimates represent the independent, direct effects of these program characteristics on program impacts. The second and third equations are discussed in the next section as part of the model specification.

---

[6] Site impact estimates that control for differences in client characteristics are often referred to as conditional impacts.

Before proceeding to this discussion, however, it is important to note that the two levels of the model are estimated simultaneously.[7] Thus, estimates of the effect on program impacts of each variable at one level control for all other variables at that level plus all variables at the other level. For example, estimates of the effect of clients' prior education on their program impacts control for all other client characteristics (in Level One) plus all features of program management, program services, and economic environment (in Level Two). Likewise, estimates of the effect of a particular management feature control for all other factors in Level Two plus all client characteristics in Level One.

2.2.3 Specification of the Model

The following series of equations specifies the model used for the present study.

LEVEL ONE

$$Y_{ji} = \alpha_j + \beta_j P_{ji} + \sum_k \delta_k CC_{kji} + \sum_k \gamma_k CC_{kji} P_{ji} + \kappa_j RA_{ji} + \varepsilon_{ji} \qquad (1)$$

where client characteristics are grand-mean centered and:

Y_{ji} = total two-year follow-up earnings for sample member i from office j,
P_{ji} = one if sample member i from office j is a program group member and zero otherwise,
CC_{kji} = client characteristic k for sample member i from office j,
RA_{ji} = a zero/one indicator variable to distinguish members of two sample cohorts at office j that were subject to different random assignment ratios,
α_j = mean two-year follow-up earnings at office j for the typical control group member from the full study sample,
β_j = the program impact at office j for the typical program group member from the full study sample,
δ_k = a regression coefficient indicating how mean two-year follow-up earnings vary with client characteristic k,
γ_k = a regression coefficient indicating how impacts vary with client characteristic k,
κ_j = the regression-adjusted difference in mean follow-up earnings for control group members in the two random assignment cohorts at office j,
ε_{ji} = a random error term for sample member i from local office j.

---

[7] This is accomplished through a combination of maximum likelihood and weighted least squares procedures (Bryk and Raudenbush, 1992).

Equation 1 specifies the outcome Y_{ji} (total earnings for the two-year post-enrollment follow-up period) for each sample member as a function of a zero/one indicator variable indicating membership in the program group or control group (random assignment status), plus a series of client characteristics, a series of interactions between random assignment status and client characteristics, and a zero/one indicator variable indicating membership in one of two random assignment cohorts at each local office.

The coefficient β_j represents the conditional program impact for office j; the coefficients δ_k represent the effects of client characteristics on control group mean outcomes; the coefficients γ_k represent the effects of client characteristics on program impacts; the coefficient κ_j represents the difference between conditional mean outcomes for the two random assignment cohorts at office j.

Because all client characteristics were grand mean centered (they were measured as deviations from their mean for the full sample of 69,399 program and control group members), the values for β_j represent the program impact for the typical member of the full study sample (the sample member with full-sample mean values for all client characteristics).[8]

LEVEL TWO

Conditional Program Impacts by Office

$$\beta_j = \beta_0 + \sum_m \pi_m PM_{mj} + \sum_n \phi_n PS_{nj} + \psi EE_j + \mu_j \tag{2}$$

where all independent variables are grand-mean centered and:

β_j = the conditional program impact at local office j for a typical program group member from the full study sample,
PM_{mj} = program management variable m for local office j,
PS_{nj} = program service variable n for local office j,
EE_j = the economic environment variable for local office j,
β_0 = the grand mean program impact for a typical program group member from the full study sample,
π_m = the effect of program management feature m on program impacts,
ϕ_n = the effect of program service n on program impacts,
ψ = the effect of the economic environment on program impacts,
μ_j = a random component of the program impact for office j.

---

[8] Bryk and Raudenbush (1992) pages 25-29 present different options for centering the variables in a hierarchical model and describe how these options affect the interpretation of their coefficients.

Equation 2 specifies the conditional program impact on mean two-year post-enrollment earnings β_j for each local program office as a function of a series of program management variables plus a series of program service variables and a local economic environment variable. The coefficients, π_m, ϕ_n, and ψ represent the corresponding effects of these program characteristics on program impacts, and are the primary coefficients of interest in the present analysis. Because all independent variables in the equation were grand mean centered (they were measured as deviations from the mean value for all 59 local offices in the sample), β_0 represents the grand mean impact for the typical sample member from the typical sample office.

Control Group Conditional Mean Outcomes by Office

$$\alpha_j = \alpha_0 + \lambda EE_j + \upsilon_j \tag{3}$$

where the economic environment variable is grand-mean centered and:

α_j = the conditional control group mean earnings at local office j for a typical member of the full study sample,
EE_j = the value of the economic environment variable for local office j,
α_0 = the grand mean conditional control group earnings for a typical member of the full study sample,
λ = the effect of the economic environment on control group earnings,
υ_j = a random component of the conditional control group mean earnings for office j.

Equation 3 specifies the conditional control group mean outcome α_j for each office (that is, control group earnings) as a function of the local economic environment. Allowing the conditional control group mean outcomes to vary across offices creates a different counterfactual for each office. This is a necessary step so that the local program office impacts (modeled in equation 2) are accurate and meaningful.

The coefficient λ in Equation 3 represents the effect of the local economic environment on control group outcomes. Because the economic environment variable is grand-mean centered, the coefficient α_0 represents the mean outcome for the typical control group member for the typical program office.

Random Assignment Cohort Outcome Differences by Office

$$\kappa_j = \kappa_0 + \eta_j \qquad (4)$$

where:

κ_j = the difference in conditional mean earnings for the two random assignment cohorts at office j,
κ_0 = the grand mean difference in conditional mean earnings for the two random assignment cohorts, and
η_j = a random component of the difference in conditional mean earnings for the two random assignment cohorts at office j.

Equation 4 specifies a simple random variation across offices in κ_j, the difference in conditional mean outcomes for their two random assignment cohorts. This variable and its equation are included in the model to represent the fact that, in some offices, the random assignment ratio (the ratio of program group members to control group members) was changed during the sample enrollment period.[9] Because the coefficient for this variable does not have an important substantive interpretation, it is not reported in the findings tables later in the paper.

2.2.4 Planned Future Extensions of the Model

The preceding model represents the first step in a more comprehensive program of research. Thus, future research will explore some important analytic issues that are not addressed by the present analysis. To do so, the current model will be extended to test for:

threshold effects that might exist if program performance responds in a nonlinear way to *extreme variations* in a particular program feature,

interaction effects that might exist if program performance responds in a nonlinear way to a *particular combination* of program features and/or to a particular combination of program features and client characteristics,

---

[9] For administrative reasons, some local offices had to change their program/control group random assignment ratio at least once during sample enrollment. Hence, their program/control group mix varies across enrollment cohorts. To reflect this in the analysis, enrollment cohorts were allowed to have different mean control group outcomes. This was accomplished by adding a zero/one random assignment cohort variable to the model. For the offices where this problem never arose, values for the variable were assigned randomly to create a placeholder because Equation 4 required all offices to have such a variable. For the few offices that changed their random assignment ratio more than once, a fraction of the sample members were randomly deleted to construct a sample with only two random assignment ratios per office. These modest accommodations to the reality of random assignment had no effect on the findings from the model.

indirect effects that might exist if program performance responds to a program feature *through its effect* on program services, and

other program effects that might become apparent if program performance is measured in terms of additional *short-run and long-run labor market and welfare outcomes*.

Chapter 5 briefly describes these potentially important effects and outlines how they will be addressed in the future. Doing so, however, is not likely to fundamentally alter the main conclusions of the present analysis, given the robustness of its findings. Instead it is hoped that future analyses will provide a more nuanced understanding of how client and program characteristics affect program performance.

Chapter 3
The Settings, Sample, and Measures

This chapter briefly describes the program settings that are the focus of the present study and the analysis sample of female single parents whose experiences are analyzed. It then describes the measures included in the analysis and the data used to create them.

3.1 The Settings

The welfare-to-work programs examined by the GAIN, PI, and NEWWS evaluations were operated as the various states' own versions of the federal Job Opportunities and Basic Skills Training (JOBS) program authorized and funded under the Family Support Act of 1988. The primary objectives of these evaluations were to: (1) determine whether the programs being studied increased client employment and earnings, and decreased their welfare receipt compared to a control group of persons who were not offered program services; (2) compare the benefits and costs of the programs; and (3) study how the programs were implemented, the problems they confronted, and how these challenges were addressed. Table 1 lists the program enrollment/random assignment period for each evaluation, the number of local program offices it included, the number of counties it included, and the state or states in which it was conducted.

3.1.1 The Local JOBS Offices

The present analysis focuses on the impacts and characteristics of 59 *local program offices* where participants in GAIN, PI, and NEWWS received their JOBS case management services. These services included, among other things, assessment of clients' service needs, development of client employability plans, assignment of clients to program activities, arrangement of client support services (child care, transportation, etc.), and monitoring of client participation in program activities. In some sites, JOBS caseworkers also conducted orientations, provided intensive client counseling, and performed specialist functions such as job development. Although caseworker roles varied within and between offices, the preceding characterization of JOBS casework applies to most offices in the GAIN, PI, and NEWWS studies (Riccio and Friedlander, 1992, p. 53; Kemple and Haimson, 1994, p. 36; Hamilton and Brock, 1994, p. 62).

The employment-related services that are the focus of caseworkers' roles in the JOBS offices are typically separate from income maintenance functions that other welfare workers perform in determining people's initial or continuing eligibility for welfare and other transfer benefits, and the amount of their grants. In some sites, however, both sets of functions were combined and assigned to the same staff. But even where they were kept separate, the staff performing those different functions had to coordinate their efforts. Hamilton and Brock (1994) distinguish as follows between this income maintenance, or "AFDC case management," and JOBS (i.e., employment-related) casework:

> JOBS and AFDC case management are usually conceptualized as two distinct roles. In most states, welfare recipients see an income maintenance worker to apply and retain eligibility for AFDC, food stamps, Medicaid, and other benefits and meet with a separate JOBS worker on all matters pertaining to their JOBS participation. Where income maintenance and JOBS functions intersect – for example, in granting exemptions to welfare recipients from the JOBS participation requirement or imposing financial sanctions for noncompliance – the two types of workers have to coordinate with each other. (Hamilton and Brock, 1994, p. 66)

The present analysis focuses on specific strategies used by JOBS caseworkers that concern their efforts to help move their clients into jobs. It develops indicators of these strategies and examines the relationship of those indicators to program performance.[1]

3.1.2 The Three Evaluations: GAIN, PI, and NEWWS

GAIN served as California's JOBS program. In contrast to earlier welfare-to-work initiatives, GAIN was noted particularly for the importance it placed on basic education for those determined to need remediation in basic reading or math skills or instruction in English as a Second Language. The program also provided job search assistance, unpaid work experience, and referrals for post-secondary education and vocational training (Riccio and Friedlander, 1992). Participation in GAIN, as in all JOBS programs, was mandatory for a large part of the welfare caseload.[2] Those who were required to participate but failed to do so without what was considered to be "good cause" were to receive a financial sanction, i.e., a penalty in the form of a reduction in the family's welfare grant. Welfare recipients who were mandated to participate in GAIN and who attended a GAIN orientation were randomly assigned primarily to either a program group or a control group; random assignment occurred at the GAIN orientation session. Appendix Figure A1 illustrates the flow of steps in the GAIN program model.

PI served as Florida's JOBS program. Although it, like GAIN, provided a range of activities and services and included a similar participation mandate, it focused particularly on low-cost job search strategies and more limited access to basic education (Kemple and Haimson, 1994). Also in contrast to GAIN, random assignment occurred at an earlier point in PI, when individuals first applied for AFDC benefits or when their benefit eligibility was determined or redetermined. Appendix Figure A2 illustrates the flow of steps in the PI program model.

---

[1] As described later, the Columbus, Ohio site in NEWWS implemented a special random assignment design to test different models of case management in JOBS offices and income maintenance offices.

[2] Mandatory GAIN recipients were originally defined as "single parents whose youngest child was six or older and the heads of two-parent households" (Riccio and Friedlander, 1992, p. 11). The single parents in this group make up most of the GAIN sample members included in the present study. However, it should be noted that when GAIN became California's JOBS program, the mandatory population was expanded to include those whose youngest child was at least 3 years of age.

Sharp cutbacks in the provision of PI services during the second half of the original study period are a distinguishing aspect of PI's implementation, which has important implications for the present analysis. These cutbacks were a response to funding reductions, a staff hiring freeze, and rapid welfare caseload growth that strained limited existing resources (Kemple, Friedlander, and Fellerath, 1995, pp. 6-9). Thus, two enrollment cohorts, which were exposed to two essentially different PI programs, were identified by the original evaluation. The "early" cohort contained sample members who were enrolled and randomly assigned between July and December 1990; the "late" cohort contained sample members who were enrolled and randomly assigned between January and August 1991. Only the late cohort was included in the present analysis because the staff survey used to measure program characteristics was administered only during the time when this group was in the program (see Appendix C).

NEWWS is a six-state evaluation of alternative mandatory welfare-to-work strategies. The programs included in this evaluation offered a range of activities similar to those offered by GAIN and PI. All NEWWS sites have operated under JOBS program rules, and, in all but two sites, random assignment was conducted at the point of orientation for the JOBS program. In two sites, however, random assignment was conducted at the point when clients applied for AFDC benefits or had their eligibility redetermined (as in PI).

Perhaps the most unique feature of NEWWS is that three of its sites—Atlanta, Georgia; Grand Rapids, Michigan; and Riverside, California—implemented a three-way random assignment design in order to permit direct comparisons of the effectiveness of labor force attachment and human capital development strategies relative to a common control group (see Appendix Table A1).

A fourth site, Columbus, Ohio, implemented a three-way random assignment design to assess different ways of organizing case management. Clients at this site were randomly assigned either to one program group for which case management and income maintenance tasks were performed by separate case managers (the "traditional" strategy); a second program group, for which case management and income maintenance tasks were performed by the same case manager (an "integrated" strategy); or a control group, which was not subject to JOBS program requirements and did not receive JOBS services. Both program groups placed a substantial emphasis on basic education and job skills training.

The three other NEWWS sites—Detroit, Michigan; Oklahoma City, Oklahoma; and Portland, Oregon—randomly assigned clients to their JOBS program or a control group. The Detroit and Oklahoma City programs were mainly education-focused, while the Portland program emphasized labor force attachment with a mix of education and training activities.

Although the programs evaluated in the original GAIN, PI, and NEWWS studies were guided by distinctive formal programmatic models or designs that applied to each of their respective local offices, it important note that management and staff practices, client participation patterns, and even local economic conditions differed substantially *across the offices* that were *within* each of the evaluations and the states and counties participating in them. This is essential to the goal of the present study, which is to understand whether variation in office-level factors is related to the variation in office-level impacts.

3.2 The Sample

Data for this analysis were obtained from administrative records, client information forms, client follow-up surveys, and staff surveys for each of the 59 program offices included in the present study. Table 2 summaries the sample sizes for each data source. For reasons discussed below, these samples are subsets of the full samples used for the GAIN, PI, and NEWWS evaluations. Hence, there may be differences between specific findings in this paper and those presented in the reports on the original studies.

3.2.1 Data Sources and Sample Sizes

The sample of clients for the present analysis includes 69,399 female single parents assigned to the program and controls groups—46,977 from NEWWS, 18,126 from GAIN, and 4,296 from PI. The combined group of control and program clients for a local program office averages 1,176 persons and ranges from as few as 177 to as many as 4,418.

Sample intake forms provide information on clients' socio-economic characteristics. Administrative records from quarterly Unemployment Insurance (UI) system records and monthly AFDC case records provide information on their earnings and AFDC receipt.

Information about program management was obtained from staff surveys (in the form of self-completed questionnaires) administered to 1,225 JOBS caseworkers at the 59 local program offices. These surveys provide information about program implementation, interactions between program staff and their clients, and the relationships between staff and their supervisors. On average, 21 staff members per office responded to the survey, with a standard deviation of 19 and a range of 1 to 83. Completion rates for these surveys were uniformly high, exceeding 90 percent in most offices.

Information about clients' receipt of program services is only available for a random subsample of clients at each site who were interviewed as part of follow-up surveys administered to the program and control groups. These surveys included special modules of questions on the use of employment-related services accessed through the welfare-to-work programs or independently of them. These data were used to measure the percentage of program group and control group members at each local office that received job search assistance, basic education, and/or vocational training. The client follow-up survey sample for a program office averaged 258 persons, with a standard deviation of 423, and a range of 27 to 2,159. Response rates to the survey ranged from 70 to 93 percent across the study counties.

3.2.2 Analysis Sample

Three restrictions were applied to derive the present analysis sample from the original evaluation samples. First, to simplify the interpretation of the analysis, only females were included. Second, to facilitate estimation of the impact model, only offices with data for all program characteristics were included.[3] Third, two GAIN offices were dropped because of their especially small samples, and a third was dropped because of its unusual client mix.[4]

3.3 The Program Performance Measure

As noted earlier, program performance was measured for the present analysis as the *estimated program impact on total client earnings for the first two years after random assignment*. Earnings data for each sample member were obtained from state UI wage records, which provide quarterly information about earnings from all jobs that are covered by unemployment insurance. Well over 95 percent of legitimate jobs (i.e., not "under-the-table" jobs) in most states are covered by the UI system.[5] When a sample member is not employed in a UI-covered job during a quarter, zero earnings are recorded for that quarter. These zeros are then included in total earnings for each sample member's two-year follow-up period.

Because the random assignment dates for sample members vary both within and across offices, two strategies were used to account for these timing differences. First, to assure that all earnings amounts are comparable over time, they were converted to constant 1996 dollars using the CPI-U (Economic Report of the President, 2000). Second, to align total earnings temporally the same way for all sample members, each client's earnings were calculated as the sum of her earnings for the first eight quarters after her quarter of random assignment.

With comparable measures of total follow-up earnings for each client in every office, it was possible to estimate the impact of each local office as a regression-adjusted difference in the mean earnings of its program and control group members. This impact represents the *value added* by the program.

---

[3] The client survey, which collected information about service receipt, was administered in two offices in one PI county, and in one office in each of the remaining eight PI counties (for a total of ten offices). Fifteen additional PI offices (which have no service receipt data) were dropped from present study. In addition, the client survey was not administered in the Butte GAIN office, so it was also dropped.

[4] About 98 percent of clients in this office were Asian, 67 percent were female, and only 15 percent had a high school degree or GED.

[5] Kornfeld and Bloom (1999) describe the types of data collected by the UI system and assess its validity for measuring earnings for low-income persons.

As a starting point for comparing office performance, Table 3 lists the impact estimates for the 59 offices in the present study. These are unconditional impacts; that is, interactions of client characteristics with the program group indicator variables are *not* included in the Level One model in this specification (see Appendix B). Each office is identified by a label indicating the evaluation in which it participated and the rank order of its impact for that evaluation. Thus, for example, office GAIN1 is the GAIN office with the most positive program impact for GAIN; NEWWS2 is the NEWWS office with the second most positive impact for NEWWS; and so forth.[6] This labeling scheme is used instead of naming each office, because the present analysis is designed to study how program characteristics affect program performance in general, not to identify "best practices" in a few select offices.

The first column in Table 3 lists the impact estimate for each office; the second column lists the standard error for each impact estimate; and the third column lists its p-value—a measure of statistical significance.

The impact estimates in the table range from a low of - $1,412 to a high of $4,217, with a mean of $883 and a standard deviation of $1,182 (all in 1996 dollars).[7] This mean impact is not insubstantial: it represents 18 percent of the average earnings of the control group. In other words, on average, the programs increased their clients' earnings by 18 percent above what they would have been in the absence of the interventions. Viewed from another perspective, the impact of $883 ranges from 6 to 13 percent of the maximum total AFDC benefits that a family of three could receive during a two-year period, depending on the state of residence. Thirteen of the impact estimates are negative, but their p-values indicate that none of them are statistically significant at conventional levels. Hence, there is little evidence that the local offices actually reduced clients' earnings.

In contrast, twenty-four of the positive impact estimates are statistically significant at the 0.10 level and twenty of these are significant at or beyond the 0.05 level. Furthermore, many of these estimates are quite large. Hence, there is substantial evidence that many of the local offices increased clients' earnings.

It is especially important for the present analysis that the *variation in impacts across offices* be substantial and statistically significant, because this variation is the basis for estimating the relationships between program characteristics and program impacts. Thus, for example, if impact estimates were statistically significant and large for every office, but all estimates were the same, there would be no variation for program characteristics to "explain," and thus, no information from which to determine how they influence program impacts. Fortunately, as shown in Table 3, program impacts do vary substantially across offices and, as documented in Appendix B, this variation is highly statistically significant.[8]

---

[6] These labels are the same for all tables that identify offices.

[7] This mean of $883 weights each site equally and is presented for descriptive purposes only. It differs from the estimated grand mean impact, which is reported later in this paper, which weights each site according to the reliability of its impact estimate.

Figure 2 provides another way to view the variation across offices. This figure also shows the differences between observed and "true" impacts, discussed in section 2.2.1. The top panel of Figure 2 is a histogram of the observed program impacts, like those reported in Table 3. These are unconditional impacts estimated by OLS. As described in section 2.2.1, the observed impacts consist of two components: (1) estimation error, and (2) "true" impact. The bottom panel of Figure 2 is a histogram of these unconditional true, underlying impacts, estimated by empirical Bayes methods in HLM (Bryk and Raudenbush 1992, pp. 39-40). These "true" impacts, with estimation error removed, will be modeled using client characteristics in Level One and program management, service receipt, and economic environment as explanatory variables in Level Two.

3.4 The Program Management Measures

As previously discussed, six program management measures are used in the present analysis: (1) the degree of a program's emphasis on rapid employment, (2) the intensity of its focus on personalized client attention, (3) the degree to which it emphasizes close monitoring of client participation, (4) the consistency of frontline staff members' perceptions about these three program features, (5) the agreement between frontline staff and their supervisors about these three program features, and (6) the size of staff caseloads. The first three measures represent how caseworkers (i.e., the frontline staff) define the core elements of the service technology for their programs. The next two measures represent the consistency of staff and supervisor views about these elements. The final measure represents the availability of perhaps the most important program resource—staff time and energy.

Because frontline staff play a central role in determining the actual nature of the intervention offered by a welfare-to-work program (Lipsky 1980), measuring the variation across offices in their perceptions and practices is one important way of capturing how offices differ along critical dimensions that managers can influence. This variation, in turn, may help explain differences in what clients actually experience in their programs and, ultimately, variation in the success of those programs in improving labor market and welfare outcomes.

The surveys administered to staff in each program office were the source of data for the six management measures previously discussed. Responses to these surveys provide information about local organizational conditions, interactions between staff and their clients, and relationships between staff and their supervisors. Both caseworkers and their unit supervisors responded to the surveys.

---

[8] A Chi-Square test was used to assess the statistical significance of the variation in impacts across offices. Bryk and Raudenbush (1992, pp. 54-56) discuss the computation and interpretation of this test statistic.

Caseworkers are the primary point of contact for most program clients, but in some cases, supervisors also see a few clients regularly or fill in for staff when they are on vacation or ill.[9] Responses from frontline staff were used to construct five of the six management measures. Responses from both frontline staff and their supervisors were used for the sixth measure of the difference in their perceptions. Appendix Section C.1.1 contains additional information about the surveys from which data for the measures were obtained.

To measure ***program emphasis on quick job entry (Scale 1)***, the present study uses a scale based on frontline staff responses to the four staff survey questions paraphrased in the first panel of Table 4. Appendix Section C.1.2 describes how these responses were combined for each staff member and how responses were further aggregated to produce a measure for each local office. To facilitate its interpretation in the analysis, the office-level measure was standardized to have a mean of zero and a standard deviation of one. As shown in Table 5, its values range across offices from a low –1.7 to a high of 2.5, with higher values indicating greater emphasis on quickly moving clients into jobs.

Program emphasis on personalized client attention (Scale 2) is measured using a scale constructed from responses to five staff survey questions (paraphrased in the second panel of Table 4) concerning how much effort staff make to learn about the client's needs and circumstances in depth and to tailor services accordingly. The construction of this scale was similar to that for the "quick job entry" scale, described above. Its values range across offices from a low of -2.0 to a high of 2.3, with higher values indicating greater program emphasis on personalized client attention.

To capture ***how closely staff monitor their clients' participation in program activities (Scale 3)***, the study uses a scale constructed from five survey questions (paraphrased in the third panel of Table 4) that pertain to the staff awareness of clients' attendance and performance problems. Values for the scale range across offices from a low of –2.8 to a high of 1.9, with higher values indicating more intensive monitoring.

To the extent that managers successfully instill a common vision within their own organizations, staff views on the key elements of their service technology should converge. Two different scales were constructed to measure the extent to which a common vision did or did not exist within an office. One of these measures the degree of ***inconsistency between staff and supervisor views of the organization's service technology (Scale 4)***. To the extent that managers effectively impart a common message, the views of staff and their supervisors about these issues should converge. A high office score on this measure reflects a high degree of disagreement between frontline staff and their supervisors. A second scale measures ***inconsistency in the frontline staff's views*** on those same dimensions of the program's service technology ***(Scale 5)***. A high office score on this scale means that the staff of that office differ widely among themselves in how much emphasis they say they place on quick employment, personalized attention, and closeness of client monitoring.

---

[9]. Available data do not indicate the extent to which supervisors interacted with clients.

Appendix Section C.1.2.3 describes how these two scales were constructed. The first one reflects the within-office difference between the frontline staff and their unit supervisors in their average responses on each scale. The second reflects the within-office variation (measured as a standard deviation) in responses among frontline staff on the three service technology scales, pooled across those scales. Both scales were constructed to have a mean of zero and a standard deviation of one. Values for the "staff/supervisor inconsistency scale" range from −1.5 to 3.2; values for the "staff inconsistency scale" range across offices from −2.1 to 4.5. If a manager's success in instilling a common vision of their program matters for program success (performance), offices that rate lower on these scales (i.e., have less disagreement) will have higher program impacts.

Finally, a measure of ***caseload size (Scale 6)*** was constructed from responses of frontline staff to the following question on the staff surveys: "How many clients are on your caseload today?" The average response to this question for each office provides a measure of the size of its average caseload. Based on these responses, the average office in the present sample had 136 clients per caseworker. The standard deviation across offices was 67 and the range was 70 to 367. In comparison to the Riverside GAIN experiment, the much higher caseloads of staff in many of the offices included in the present study makes it possible to test whether a much wider variation in caseload size matters for program performance.

3.5 The Program-Induced Service Differential Measures

This section describes how three measures of the difference in rates of participation in employment-related activities between the program and control groups—referred to here as *service differential measures*—were constructed.

A random subsample of program and control group members from each local office was interviewed roughly two years after their random assignment date as part of follow-up surveys for the original GAIN, PI, and NEWWS evaluations. Among other information collected by the surveys, respondents were asked about their receipt of specific program services during the preceding two years. The present analysis uses this information to characterize the program-induced service differential for each local office.

To define these measures, it was first necessary to make two key decisions. The first decision concerned the *types of services* to include. Three core types of services were examined in the present analysis: job search assistance, basic education, and vocational training. Each of these categories encompasses a broad range of specific activities: *job search assistance* includes both individual job search efforts and group-oriented job club activities; *basic education* includes adult basic education classes, GED preparation courses, and classes in English as a second language (ESL); *vocational training* includes classroom training, on-the-job training, unpaid work experience, and post-secondary or vocational training. Clients may participate in any number or combination of these activities.

The second major decision concerned how best to *measure service receipt.* For this decision there were three basic alternatives: (1) measuring the services received by program group members only, (2) measuring the services received by control group members only, or (3) measuring the difference between the services received by the two groups. This latter approach was chosen because it provides an estimate of the difference between services actually received by program group members and what they would have received without the program being tested. In other words it represents their *program-induced service differential*, which, in turn, should be directly related to their program-induced earnings gain (i.e., impact).

As background information, the first row in Table 6 lists the average percentage of program group members who received each of the three basic types of services. It shows that 19 percent received basic education, 22 percent received job search assistance, and 27 percent received vocational training. (Some clients entered these activities on their own during the follow-up period but after exiting the program or welfare.) The second row in the table lists the average percentage of control group members who received each type of service: 8 percent, 5 percent, and 22 percent, respectively. The third row lists the average difference between program and control group receipt rates (i.e., their average program-induced service differential): 11 percentage points for basic education, 17 percentage points for job search assistance, and 5 percentage points for vocational training.

As can be seen, even though the largest percentage of program group members received vocational training, its service differential was the smallest because it was the most popular self-initiated service (control group members were especially likely to obtain it on their own). At the other extreme, the program-induced service differential was greatest for job search assistance, the activity that members of the control group were least likely to enter on their own.[10]

The averages mask a substantial variation across offices. Summary measures of this variation are listed in the last two rows of Table 6 and in the middle panel of Table 5, and values for each individual office are provided in Appendix Table C7. The cross-office standard deviation was 13 percentage points for basic education, 12 points for job search assistance, and 10 points for vocational training. This reflects a very broad range from negative double-digit values (indicating service receipt rates that are higher for control group members than for program group members) to positive values in the vicinity of 50 percentage points. Although some of this variation represents estimation error, statistical tests indicate that the variation in

---

[10] It likely, however, that a number of survey respondents forgot some of the independent job search efforts they made or did not consider it as part of a program when they were asked about it during the client follow-up survey. Thus, its receipt rate is probably understated for both program group and control group members.

"true" service differentials is highly statistically significant. It is this variation that makes it possible to relate program-induced service differentials to program-induced earnings gains.

3.6 The Measure of Economic Environment

The importance of the economic environment for the success of welfare-to-work programs has been the subject of speculation for decades. However, little is known about this potential connection. To explore it, the present study uses the prevailing county unemployment rate for each program office to measure the condition of its labor market environment. Unemployment data were obtained from monthly, county-level statistics provided by the U.S. Bureau of Labor Statistics, Local Area Unemployment Statistics; and the California Employment Development Department.

Because random assignment dates vary within and across offices, the follow-up period—and, hence, the relevant unemployment rate—varies across individuals within an office as well as across offices. To account for this when constructing the office-level measure of average unemployment, an average rate was first computed for the two-year follow-up period for each client. Client averages were then aggregated to office averages. Appendix Section C.3 provides further details about how this measure was constructed.

The average unemployment rate for the present sample of offices was 7.4 percent, with a standard deviation of 3.1 and a range of 3.5 to 14.3 (see the last row in Table 5). Appendix Table C8 lists the average unemployment rate for each office, plus the within-office standard deviation, and the range of average unemployment rates faced by individual clients.[11]

3.7 The Measures of Client Characteristics

As noted earlier, including client characteristics in the present analysis serves two purposes: (1) it provides information about how program impacts differ for different kinds of clients, and (2) it controls for observable differences in the client mix across offices (and thus potential differences in the difficulty of serving their clients) when examining the relationships between program characteristics and program impacts.

Data on client characteristics were obtained from background information forms completed during enrollment of the 69,399 sample members. These forms provide information about their educational background, welfare history, number of children, and other personal characteristics.

---

[11] An additional measure of the local economic environment – county-level job growth during the two-year follow-up period – was considered and constructed, but not included in the present analysis. Appendix Section C.3.3 describes the construction of this measure and explains that it was not used in the analysis due to concerns about its likely measurement error.

Table 7 lists the individual client characteristics used for the present analysis. All characteristics are specified as indicator variables with a value of zero when the characteristic does not apply to a sample member and one when it does. The first column in Table 7 lists the percentage of clients with each characteristic in the full sample.

To describe how the client mix varies across offices, the second column in the table lists the cross-office standard deviation of the *percentage of clients in each office* having each characteristic and the third column lists the corresponding minimum and maximum percentages.

Putting these pieces together, first note that the top line in the table indicates that 56 percent of all sample members had at least a high school degree or GED prior to random assignment. The percentage of clients with this level of education varies across offices, however from a low of 17 percent in one office to a high of 74 percent in another. The cross-office standard deviation was 14 percentage points.

In addition, note that over half of the sample members had more than one child, and 46 percent had a child under six years old; 44 percent had received welfare in all 12 months of the year before random assignment, and over half had no earnings during that year. However, these and the other characteristics in the table vary substantially across offices. Thus, it was important to control for them in the analysis.

Chapter 4
The Findings

This chapter presents results obtained by estimating the model introduced in Chapter 2 from the data described in Chapter 3. To facilitate the interpretation of these results, they are presented a number of different ways.

4.1 The Hypotheses Tested

Embedded in the present model are a series of hypotheses based on theory, past research, and professional judgment. Although expectations differ about how the factors covered by these hypotheses are related to program performance (as discussed in Chapter 2), each hypothesis is stated below in a particular way to establish a basis for testing it (not to take a position on its expected outcome). The study thus examines whether each of the following factors independently improves program performance, other things being equal:

With Respect to Management Choices:

- A stronger emphasis on quick employment
- A greater emphasis on personalized client attention
- Closer monitoring of client participation
- Smaller staff caseloads
- Less inconsistency in frontline staff practices and views
- Less disagreement between frontline staff and their supervisors

With Respect to Program Services:

- Less use of basic education

With Respect to Economic Conditions:

- Lower unemployment rates

The analysis also tests whether the program impacts are larger for certain types of clients. Specifically:

With Respect to Client Characteristics:

- Clients with moderately severe barriers to employment.

The following sections present the results of the analyses conducted to test each of these hypotheses. When assessing these findings, it will be helpful to compare the effects of a change on each key independent variable to the ***average program impact on earnings*** for the ***average sample member*** at the ***average program office,*** which is $879, or 18 percent of the average counterfactual (see bottom panel of Table 8). In other words, the average program increased the average client's earnings by 18 percent above what those earnings would have been without the program.

Before proceeding, it is interesting to note that client characteristics explain 16 percent of the variation in performance of local program offices. Together, client characteristics and program characteristics explain 80 percent of this variation in performance (see Appendix section B.5).

4.2 Management and Performance

Table 8 presents findings that test the hypotheses concerning management choices and practices.

4.2.1 Client Employment Emphasis: Take a Job Quickly

A greater emphasis on rapid employment has come to be seen by many as contributing to a more successful welfare-to-work program. Consistent with this hypothesis, the estimated positive regression coefficient for the quick job entry scale is *the largest and most statistically significant coefficient in the model.* This coefficient indicates that a one-unit increase in emphasis on quick job entry (which is measured as a multi-item survey scale with a standard deviation equal to one) increases program impacts by $720, holding all other variables in both levels of the model constant (i.e., "other things equal"). This implies a $720 *increase in impact* for a one standard deviation increase in a program's rating on the quick employment scale. Put differently, because the average program's impact is $879, this change on the quick employment scale would raise that impact to $1,599 (i.e., $879 + $720). In percentage terms, this represents an increase in the earnings impact from 18 percent of the counterfactual for the average program to 32 percent—a very large effect when compared to the typical effects measured by randomized experiments conducted on welfare-to-work programs.

Another way to present this finding (which is more relevant for some of the other variables in the model) is to report it as a partially standardized regression coefficient (column two in Table 8). This simple transformation of the original coefficient represents the projected amount (in 1996 dollars) by which impacts would change if a program's rating on the quick job entry scale were increased by one standard deviation, other things being equal.[1] Because the quick job entry scale was constructed to have a standard deviation of one, its partially standardized coefficient is the same as its original coefficient.

The third column in the table lists the p-value for each regression coefficient (both original and partially standardized), which is a measure of its statistical significance. This measure helps to guard against concluding that an estimated coefficient represents a true relationship or effect when instead it was produced by random error. The smaller the p-value is, the less likely the estimated coefficient is to represent only random error, and the more likely it is to represent a true relationship or effect (i.e., the more statistically significant it is). A p-value of 0.05 or smaller is the conventional criterion for judging a finding to be statistically significant, however, recent practice has been to accept 0.10 for this purpose; the present analysis adopts this latter standard. The estimated effect of client employment focus on program impacts is highly statistically significant, with a p-value of 0.000 (the precise value is $p = 2 \times 10^{-6}$).

---

[1] The partially standardized regression coefficient equals the original coefficient multiplied by the standard deviation of the independent variable that it represents.

The final columns in Table 8 report findings as "conditional impact intervals" in dollars (column four) and as a percentage of the average counterfactual (column five). This convention (developed for the present study) illustrates how projected program impacts vary when the value of one program characteristic spans its inter-quartile range (that is, when the value of that characteristic changes from one that is at the 25th percentile to one that is at the 75th percentile) and all other variables remain at their mean values.[2] Thus, it represents the *conditional* response of program impacts to a *standardized* change in a program characteristic for the *average* sample member at the *average* program office. For the quick employment scale, this interval is $397 to $1,361, or 8.1 percent to 27.9 percent of the average counterfactual, which is a very large difference in impacts for a two-quartile (50 percentile) difference in the value of the scale.

To test the robustness of this finding, Appendix D reports the results of sensitivity tests that selectively delete program offices from the sample and re-estimate the full model. The first set of tests deletes the four Riverside GAIN offices (because Riverside was the most successful GAIN site by far), then deletes the seven Portland NEWWS offices (because Portland was the most successful NEWWS site by far) and lastly, deletes all 11 of these Riverside and Portland offices. Even after deleting all of these unusually successful programs, the estimated regression coefficient for the quick employment scale was $525 and its p-value was 0.004.

A second series of tests was conducted by first deleting the four program offices with the two most positive and negative impact estimates, then the six offices with the three most positive and negative estimates, up through the ten offices with the five most positive and negative impact estimates. Variants of this approach are often referred to as "trimming" the data or omitting outliers. Once again, the basic finding for the quick employment scale was quite robust. Even with all ten of the office outliers omitted, its estimated regression coefficient was $399 and its p-value was 0.011.

A third series of sensitivity tests was conducted by deleting the 17 program offices (ten from PI and seven from NEWWS) that administered random assignment early in the sample intake process—at the point of welfare application or redetermination—instead of later in the process when sample members attended a GAIN, PI, or NEWWS program orientation (which was the point of random assignment for the other 42 offices). Because these 17 offices administered random assignment early during intake, there was a greater margin for fall-off between random assignment to the program group and program participation. Hence, this experimental design may have diluted the program and control group treatment contrast and thereby produced program impact estimates that were systematically different from those of the other offices. Nevertheless, when the 17 offices were omitted from the sample, and the impact model was re-estimated, the coefficient for client employment focus was $455 and its p-value was 0.010 (Appendix Table D3). Thus, the finding is quite robust to the deletion of these offices.

---

[2] To ensure consistent treatment of each independent variable, this calculation proceeds *as if* they were distributed normally across offices and sets the lower value of the inter-quartile range to 0.67 standard deviations *below* the mean and the upper value to 0.67 standard deviations *above* the mean. This represents the 25th and 75th percentile values, respectively, for a normal distribution.

A final series of sensitivity tests was conducted by deleting a progressively increasing number of offices with the most extreme positive and negative values on selected explanatory variables: those with the largest magnitude and the highest statistical significance of the program characteristics reported in Table 8. The scale for "emphasis on quick client employment" was one of these variables, and as further described in Appendix Section D.1.4, the estimate reported in Table 8 is robust to the deletion of outlier offices measured by this scale.

These findings consistently point to the same conclusion—that a strong employment message to clients that encourages them to move into the labor market quickly can be a powerful medium for increasing a program's success in raising their earnings.

4.2.2 Personalized Client Attention: Getting "Close to the Customer" Can Make a Difference

Findings for personalized client attention are also striking, statistically significant, and robust. The regression coefficient for this variable suggests that increasing it by one standard deviation will increase program impacts by $428, other things being equal. The p-value for this coefficient is 0.0002, which indicates that it is very unlikely to represent only random error. The conditional impact interval for this variable, $592 to $1,166 (or 12.2 percent to 23.9 percent of the counterfactual) indicates that a two-quartile change in its value can lead to a large change in program impacts. Lastly, sensitivity tests indicate that the regression coefficient for this variable is $334 (p-value = 0.016) without the 11 Riverside GAIN and Portland NEWWS offices; $267 (p-value = 0.011) without the ten office outliers; $204 (p-value = 0.176) without the offices that conducted early random assignment; and $441 (p = 0.008) without the offices with the highest and lowest values on this independent variable. Hence, the basic finding of a positive relationship between emphasis on personalized client attention and program success is relatively robust.

This finding indicates that the well-worn private sector adage about "getting close to the customer" may also apply to human service programs run by government agencies and not-for-profit organizations.

4.2.3 Closeness of Client Monitoring: Information Alone is not Enough

Knowing in a timely way how well clients are attending and progressing in their assigned activities is presumably essential if frontline staff are to enforce a participation mandate more rigorously, or to provide clients with more helpful case management and guidance through the program. It is therefore surprising that this study finds that, all else being equal, offices that more closely monitor clients tend to have *smaller* impacts on earnings. However, this relationship is not statistically robust.

The regression coefficient for this variable indicates that increasing the average office's rating on the monitoring scale by one standard deviation will *reduce* its earnings impact by $197, other things being equal. This implies a conditional impact interval of $1,011 to $747, or 20.8 percent to 15.3 percent of the counterfactual. However, the underlying coefficient estimate just misses being statistically significant (its p-value is 0.110).

Furthermore, sensitivity tests of the coefficient estimate produced mixed results. On the one hand, deleting the 11 program offices from Riverside GAIN and Portland NEWWS had little effect on the estimate's magnitude (which was - $173 after the deletions), but by reducing the sample size, these deletions reduced its statistical significance (to a p-value of 0.231). On the other hand, deleting the 10 office outliers, or alternatively the offices with early random assignment, caused both the magnitude and statistical significance of the estimate to erode substantially (although it remained negative).

In interpreting this finding, it is important to understand that the monitoring variable used in this study mostly measures the timeliness of staff knowledge or awareness of clients' participation patterns; it does not directly measure their efforts to enforce compliance or to provide helpful assistance to facilitate that participation. For example, theoretically, offices that took *either* a very tough *or* a very lenient stance toward enforcement could have rated high on this scale. Perhaps what matters is not just staff awareness of participation problems and non-compliance, but what staff do with the information they have on clients' participation. Awareness by itself is not enough, and those offices that most closely monitor their clients may not also be places that take the most productive steps to deal with participation problems that are detected.

4.2.4 Caseload Size: Human Resources Matter

The estimated effect of caseload size on program impacts is large, statistically significant, and robust. For example, the regression coefficient for this variable indicates that program impacts on earnings decline, on average, by $4 per additional client per caseworker, other things equal. The p-value for this coefficient of 0.003 indicates that it is highly statistically significant.

To interpret this result, however, it is more helpful to view it through the lens of the partially standardized regression coefficient. This parameter implies that increasing the caseload size by one standard deviation (67 clients) will reduce program impacts by $268, which is a sizable reduction. This impression is further reinforced by the broad conditional impact interval of $1,058 to $700, or 21.7 percent to 14.4 percent of the counterfactual.

Sensitivity tests indicate that this finding is quite robust. Throughout all of these tests, the coefficient for caseload size remains negative and, in most cases, is statistically significant.

Thus, it appears that the allocation of the principal human resource on the program side of a welfare-to-work intervention—frontline staff—matters a great deal to its success. While in accord with conventional wisdom, this finding conflicts with results of the Riverside GAIN caseload experiment (Riccio, Friedlander, and Freedman, 1994), which found no difference in earnings impacts between sample members randomly assigned to staff with a standard caseload (averaging 97 clients per caseworker) and those assigned to staff with a reduced caseload (averaging 53 clients per caseworker).

However, the present analysis examines caseloads that typically are much larger and vary much more than those for Riverside GAIN. The mean caseload size for a program office in the present study is 136 and its standard deviation across offices is 67. Thus, plus-or-minus one standard deviation from the mean spans a range from 69 clients per caseworker to 203 clients per caseworker. It therefore seems reasonable that program impacts would erode substantially when caseloads begin to approach the higher end of this range, where staff may have very limited time to devote to each client.

4.2.5 Consistency Within the Office: Mixed Results

Findings are mixed for the last two management variables— the scale measuring staff/supervisor disagreement about key elements of their service technology and the scale measuring inconsistency in views among frontline staff. These variables (although formulated in the negative) represent the extent to which program managers are able to instill a common vision of what needs to be done for their clients and how best to accomplish this task.

Findings for the staff/supervisor disagreement scale are statistically significant and consistent with the hypothesis that a common organizational vision can improve organizational performance. The regression coefficient for this variable implies that as staff/supervisor *disagreement declines* by one standard deviation, *impacts increase* by $159, other things equal. This coefficient is statistically significant (with a p-value of 0.102) and relatively impervious to sensitivity tests, except in the case when offices that conducted early random assignment are deleted.

In contrast, no statistically significant relationship is found between an increase in the degree of inconsistency in views among frontline staff and office impacts on earnings. Thus, these results appear to challenge the hypothesized importance of the management imperative to instill a common sense of mission and method. However, one possible explanation for this finding is that the "content" of that mission might be crucial. For example, office performance may be enhanced, at least in the short run, not by cultivating a highly shared vision stressing education first over quick employment (which may even harm short-run performance), but, rather, by cultivating a highly shared vision emphasizing quick job entry over education first. Thus, it may not be shared vision per se that matters, but, rather, shared vision built around an effective service technology.

4.3 Services and Performance: Increased Reliance on Basic Education Reduces Short-Run Effects

Although many observers would agree that the best way to increase the earnings of welfare recipients in the short run is to help them find a job, views differ sharply regarding expectations for the long run (see Chapter 2). The labor force attachment approach, which advocates the use of job search assistance as an upfront program activity, argues that even for the long run, finding a job quickly is the most effective strategy because it gets clients into the workplace where they can learn on the job what they most need to know to be successful. The human capital development approach argues that only by first imparting new skills and knowledge to clients through upfront formal education and training can a program prepare clients to obtain the kinds of jobs that will move them toward economic self-sufficiency.

To assess these different perspectives, three *service differential* variables were included in the program impact model to represent the program-control group *difference* in the percentage of sample members who received basic education, job search assistance, and vocational training. Service differential estimates were used instead of absolute service levels to account for the large differences that exist across program offices and by type of service in the degree to which control group members obtain similar services on their own. As previously noted, the service differential is lowest for vocational skills training. Although a high proportion of program group members took part in such activities (often on their own initiative), a nearly as high proportion of control group members did so as well, but without any assistance from the program.

Findings in Table 8 are consistent with the hypothesis that basic education will produce smaller-than-average short-run earnings gains. The regression coefficient for this variable, which is statistically significant (p-value = 0.017), implies that program impacts decline by $16 for each one-point increase in the program-induced change in the percentage of clients who receive basic education, other things being equal. Another way to state this finding (based on the partially standardized regression coefficient) is that program impacts will decline by $205 if the program-induced change in the percentage of clients who receive basic education increases by one standard deviation (13 percentage points).

For another way to understand what this finding means, note again that the overall average impact of $879 is defined for the average value for all variables in the model (i.e. for the average client at the average program office). If all other factors stayed the same, but the basic education differential was one standard deviation above its current mean (24 percentage points instead of 11 points), the program impact would be $674 instead of $879 (or 13.8 percent of the counterfactual instead of 18.0 percent). This sizable difference illustrates the reduced short-run impacts when emphasizing basic education as a program strategy.

Although these short-run effects of basic education are consistent with prior expectations, it is not clear why vocational training does not depress impacts similarly, since both program strategies require time in the classroom and time out of the workplace. One potentially important difference between these two activities, however, is that basic education need not (and often does not) have an employment focus or a direct connection to the world of work, whereas vocational training usually has both. In addition, in many of the programs, clients were often required to attend basic education as their initial activity, or strongly pushed in that direction, whether they wished to participate in that activity or not (and many did not).[3] Thus, it is possible that basic education, when incorporated into a program in this fashion, does not simply delay clients' attachment to the job market but also imposes an opportunity cost of the time clients spend attending and preparing for class. A more selective use of basic education may be more productive. (Recall that the most effective programs, such as the Riverside and Portland programs, did include some basic education among their service offerings).

Thus, it would appear that more explicit employment-focused activities are the key to success in the short run for clients of welfare-to-work programs. The question remains, however, as to whether this key will open the door to their future economic self-sufficiency.

4.4 Economic Environment and Performance: It's Harder to Increase Earnings When Jobs are Tougher to Find

Although it seems commonsensical that the economic environment of a welfare-to-work program must affect its success, there are two opposing views about the direction this effect takes. One view posits that lower unemployment rates, which imply relatively more job openings, make it easier for programs to help clients whom they are preparing for find jobs. Thus, program performance should be stronger in such environments. The other view posits that lower unemployment rates make it easier for employable welfare recipients to find jobs on their own. In other words, when demand for workers is high, recipients need less help from a program to land a job, thereby limiting the value added by participating in a program. A good economy may also leave a harder-to-employ segment of the welfare population on the rolls, which, in turn, might make it more difficult for a program with limited resources to have an effect.

Fortunately, the settings covered by the present analysis span a wide range of economic environments due to variation in time and geography. Sample enrollment took place for GAIN in California from 1988 to 1990; for PI in Florida during 1991; and for NEWWS in California, Georgia, Michigan, Ohio, Oklahoma, and Oregon from 1991 to 1994. As documented in Section 3.6, these data provide a strong empirical base for studying the relationship between unemployment rates and program impacts on earnings.

---

[3] See Hamilton and Brock, 1994.

The regression coefficient for the unemployment rate in the present model, which is highly statistically significant (p-value = 0.004), implies that a one percentage point increase in the unemployment rate reduces program impacts by $94, other things equal. Thus, if the average program in the sample experienced an increase of one standard deviation in its unemployment rate (3.1 percentage points), the overall average impact of $879 would fall by $291 to $588 (to 12.1 percent of the counterfactual). This sizable estimated decline is quite robust and withstood almost all of the sensitivity tests reported in Appendix D.

Thus, it appears that other things being equal, the performance of welfare-to-work programs will decline when unemployment rates rise and jobs for clients become harder to find. This result has particular relevance for what might be expected during periods when the U.S. economy is on the downside of the business cycle.

4.5 Client Characteristics and Performance

There is considerable interest in how the performance of welfare-to-work programs differs for different types of clients. This interest stems from desires to achieve both the equity and efficiency goals of these programs. In this regard, most attention has been focused on how program performance varies with clients' employability, which is typically measured in terms of their level of formal education, amount of prior employment experience, and extent of prior welfare dependence.

Although the connections between these characteristics and the *level* of future economic success is well established (future employment and earnings go up and welfare receipt goes down consistently as education and past employment experience go up and prior welfare dependence goes down), their linkages to program *impacts* are far less clear. As noted in Chapter 2, the two most relevant past studies of this issue come to different conclusions, although they use evaluation findings from different kinds of welfare-to-work interventions. Friedlander (1988) concludes that program impacts on earnings are largest for welfare clients in the middle range of a distribution of background characteristics related to employability, whereas Michalopoulos, Schwartz and Adams-Ciardullo (2001) do not find a clear pattern of differences across such subgroups.

What is most distinctive about these two studies is that: (1) they were able to compare the background characteristics of welfare recipients to valid estimates of program impacts, and (2) they formulated their research questions in terms of client *subgroups* defined mainly by specific categories of a single client characteristic (e.g. education level).[4] Hence, they observed how program impacts varied when a single characteristic was varied. However, they did not focus on how impacts covary with one characteristic, holding the others constant.

The analysis in this section shares the first feature of these studies. It, too, compares client characteristics to valid estimates of program impacts. However it departs in a fundamental way from the second feature by not formulating the analysis in terms of client subgroups. Instead it formulates the analysis in terms of *conditional impact variation*, or how program impacts covary with each client characteristic, holding all others constant. In fact, given the structure of the hierarchical model used to produce these findings, the conditional impact variation reported not only holds all other client characteristics in the model constant, it also holds constant the characteristics of program management, program services, and the economic environment that are included in the model.

---

[4] They also defined subgroups in terms of specific combinations of categories for several characteristics.

The findings of the relationships between client characteristics and program impacts are presented in Table 9. Because client characteristics are defined as simple categories (represented by zero/one indicator variables in the model), it is only necessary to report the regression coefficient for each variable and its p-value. These coefficients represent the regression-adjusted *difference* between (1) the program impact for the average sample member at the average program office with the specified category for a given characteristic, and (2) the impact for a person who is the same in all ways captured by the model except for one: she belongs to the "left-out" category for that characteristic (the category that is not represented by an indicator variable in the model).

Thus, for example, the regression coefficient of $653 for clients with at least a high school diploma or GED implies that the impact for sample members with this level of education is $653 *greater on average than the impact* for clients who are the same in all other ways (and participated in programs that were the same in all measured ways) except they had not attained this level of education prior to random assignment. This finding is highly statistically significant (p-value = 0.001). In other words, although people who entered the programs *without* already having a high school diploma or GED benefit from the program, those who entered *with* those credentials *benefited more*, all else being equal.[5]

For characteristics represented by more than two categories, it is simple to extend the interpretation of the regression coefficient in the table. For example, consider the findings for number of children. The regression coefficient for "had two children" (which is not statistically significant) estimates that the program impact for clients in this category is $301 greater than clients who are the same in all other ways, except that they had one or no children (the left-out category for this characteristic). In addition, the regression coefficient for "had three or more children" (which is highly statistically significant) indicates that the program impact for clients in this category is $591 greater than for clients who are the same in all other ways, except that they have one or no children.

Note that the only other statistically significant individual-level coefficient in the model suggests that, other things being equal, women who had received welfare for all 12 months during the year before they enrolled in the study sample (which may make them a more disadvantaged group) experienced program-induced earnings gains that were $444 larger than the gain experienced by women who had not received welfare this consistently but were the same in all other respects (and were exposed to programs that were the same in all respects accounted for by the model).

---

[5] Findings for the typical sample person with and without a high school degree or GED were simulated using the results of the model displayed in Table 9.

Taken together, the findings on client characteristics present a mixed picture. Program impacts are not *consistently* larger—or smaller—for individuals having characteristics that might be seen as making them easier or more difficult to employ. Thus, these results do not provide evidence in support of efforts to target these kinds of welfare-to-work programs for clients with selected characteristics.

Chapter 5
Implications and Planned Extensions of the Analysis

This chapter summarizes the main implications of the present analysis for the design and implementation of welfare-to-work programs, in particular, and for the design and conduct of research on the performance of human service programs, in general. The chapter concludes by outlining several future extensions of the present analysis.

5.1 Implications of the Present Analysis

The findings outlined in Chapter 4 suggest that:

1. **Management choices and practices with respect to how welfare-to-work programs are implemented and how their resources are deployed matter a great deal to program success**. In particular:
   - *A strong employment message can be a powerful medium for stimulating clients to find jobs.* The present findings indicate that programs that aggressively promote quick employment increase client earnings by considerably more than programs that are less aggressive in this regard. Thus, programs can be more effective when managers make this message a central staff priority.
   - *A clear staff focus on providing personalized attention to the needs, desires, abilities, and limitations of their clients can markedly increase program success.* The present findings indicate that programs that emphasize personalized client attention are more effective than those that do not. Thus, programs can be more effective when managers instill in their staffs a firm conviction that "one size does not fit all" and that "getting close to the customer" is very important.
   - *Especially large caseloads limit the time available for staff members to work directly with their clients and thereby can undermine program effectiveness.* Current and past research indicates that while especially small caseloads may not improve program performance, especially large caseloads can hurt performance substantially.
2. **Increased reliance on basic education reduces short-run effects.** The present findings indicate that offices that rely more on basic education have smaller impacts. Programs emphasizing more explicitly employment-focused activities appear to be more successful in the short run. However, the long-run payoffs of such strategies remain to be determined.
3. **The local economic environment is a major factor in the determination of program success.** The present findings indicate that program-induced earnings gains are larger when local unemployment rates are lower. This

underscores the importance of developing contingency plans that will help welfare administrators anticipate the needs that will arise when the US economy weakens.

4. **Program effectiveness for different types of clients does not follow a clear pattern.** The present findings indicate that welfare-to-work programs can be effective for many different types of clients—not just a few isolated subgroups. Furthermore, they suggest that no clear overall pattern exists in the relationships between client characteristics and program impacts. Combined with similarly mixed findings from past research, this suggests that a system of socio-economic "profiling," which attempts to identify types of clients most likely to benefit from a program and made the target of its efforts may be very difficult to construct.

5. **A systematic multi-site strategy that uses natural variation across sites to compare valid and reliable measures of program characteristics with experimental estimates of their impacts can provide important insights into the linkages between program implementation and program performance.** The present analysis illustrates the types of lessons that can be learned from this approach and, thus, helps to demonstrate its potential for future policy research. It is therefore hoped that by carefully laying the framework for such comparative analyses through the design of future experiments, researchers can begin to open the "black box" of human service programs and thereby increase their ability to provide practical solutions to important real-world problems.

5.2 Planned Extensions of the Present Analysis

As noted earlier, the present analysis is the first step in a more comprehensive program of research that will use the current approach and data to study the determinants of effective welfare-to-work programs. Hence, the present analysis does not address several potentially important issues that will be explored through future research.

5.2.1 Issues to be Addressed

Two issues not addressed by the present impact model stem from its specification as a series of linear additive functions (Equations 1 through 4 in Chapter 2). This specification was used to simplify the initial analysis and maximize its statistical power by minimizing the number of parameters to be estimated. Nevertheless, the particularly simple and mathematically convenient specification used may not fully reflect the subtleties of how program characteristics are related to program impacts in practice.

One implication of the present specification is that the change in program impacts per unit change in each program characteristic is assumed to be constant for all values of that characteristic. A second implication is that the effect of each program characteristic is assumed to be independent of the value of the others. While these assumptions may be reasonable approximations for gauging the implications of small changes in program characteristics, they may not be appropriate for predicting behavioral responses to large changes.[1]

Another way to think about these simplifications is to note that they imply the absence of *threshold effects* and *interaction effects.* Threshold effects, which in some contexts are referred to as "tipping,"[2] can occur when a program characteristic does not affect program impacts until it exceeds or drops below a certain threshold level, beyond which impacts change precipitously. For example, it is possible that unemployment rates do not affect program impacts until they approach a very low level, at which point impacts may increase substantially. Similarly, it is possible that staff caseload size does not affect program impacts until it exceeds a particularly high level, beyond which program impacts drop sharply.

Interactions reflect synergies among program characteristics that can occur when a specific combination of them has a pronounced effect that exceeds the sum of their separate effects.[3] For example, vocational training provided in the context of limited emphasis on quick client employment and/or limited job search assistance might have no effect on program impacts, whereas it might have a substantial effect when these other factors are present to a greater degree and thereby help to convert new skills into new jobs. Interactions can also occur between client characteristics and program strategies. It may be that certain program strategies are very effective with clients having a broad range of characteristics, while other program strategies work well for only certain types of clients.

A third issue for future research is the possibility that program services may be *intervening variables* that mediate the relationships between other program characteristics and program impacts. In other words, the present model does not account for the possibility that management practices, the economic environment, and client characteristics have *indirect effects* on program impacts through their effects on what program services are received by sample members, and the subsequent effects of these services on program impacts. Instead, the model specifies only the *direct effects* of program characteristics (presumably through their influence on how employment mandates promote client employment). A more complete model would specify both direct and indirect effects, which would make it possible to estimate the *total effect* of each program characteristic.

A fourth issue for future research is the relationship between program characteristics and program impacts on other labor market and welfare outcomes, both in the short run and in the long run. The present analysis focuses on one short-run outcome—total earnings during the first two years after random assignment. However, there is reason to believe that different types of outcomes may respond differently to the same program characteristic. Furthermore, as noted earlier, there is reason to believe that impacts in the long run may differ from those in the short run. Thus, the present analysis represents only one piece of a potentially much larger puzzle.

---

[1] The present model is equivalent to approximating a nonlinear function with a first-order Taylor Series (Greene, 1997, pp. 452-453).

[2] The best-known example of tipping involves racial transitions in neighborhoods. This concept applies to a broad range of other phenomena, however (Gladwell, 2000).

[3] More generally, interactions represent any situation where the sum of the separate effects of a group of variables is less than or greater than their combined effect. In the present context, such interactions might involve two or more Level One client characteristics, or two or more Level Two program characteristics, or a combination of program characteristics and client characteristics.

5.2.2 How These Issues can be Addressed

Future research will address these outstanding issues by extending the present analysis. For example, data are available for a number of additional labor market and welfare outcomes defined for the current two-year follow-up period and Table 10 illustrates their potential for expanding the scope of the present study.

The top panel of the table reports impact information for three additional two-year follow-up outcomes: the mean number of quarters employed during this period, the mean total amount of AFDC payments received, and the mean number of quarters for which AFDC benefits were received. The first column in the table indicates the overall average impact on each outcome and the second column reports these impacts as a percentage of their counterfactual. Thus, for example, the overall average impact on mean quarters employed during the two-year follow-up period was 0.36 quarters, which was 15.7 percent of its counterfactual (the number of quarters that would have been employed in the absence of the programs). The third column in the table, which reports the statistical significance of each mean impact, indicates that all of the impact estimates are highly significant.

More important, however, are the remaining columns, which summarize the *variation* in impact estimates across local offices and thereby illustrate the potential for studying the effects of client and program characteristics on these impacts. Column four reports the range of estimated impacts across offices. These estimates do not control for differences in client characteristics and thus are unconditional. Column five indicates that unconditional impacts vary highly significantly across offices. Hence, there is variation to be explained by differences in client characteristics, program factors, and/or local economic conditions. The final column in the table reports the statistical significance of the variation in conditional impacts across offices, which control for client characteristics. As can be seen, even after applying these controls, there is still real variation to be explained by program factors and local economic differences.

The bottom panel in the table reports corresponding information for program impacts defined in terms of the labor market and welfare experiences of sample members during the eighth quarter after their quarter of random assignment. This information can be used to assess the extent to which program impacts are sustained during the first two follow-up years. As can be seen, these impacts are still fairly large, ranging, on average, from 10.6 percent to 15.7 percent of their underlying counterfactuals. Of greater importance, however, is the fact that the variation in both unconditional and conditional impacts across offices is still highly statistically significant. Hence, an opportunity is available to study how program and client characteristics and local economic conditions influence these dimensions of program performance.

In addition to the preceding two-year follow-up measures, corresponding data for a five-year follow-up period are currently available for GAIN and soon will be available for NEWWS. Thus, future analyses will be able to explore how program and client characteristics affect longer-run impacts. As noted earlier, this issue is especially important for providing guidance about the most effective mix of program services.

In addition to expanding the range of outcome measures considered, it will also be possible to explore the effects of program characteristics that were not included in the current analysis. One of the constraints on this option, however, is that not all staff survey questions (the basis for measuring management characteristics) were asked for all three of the studies used for the current analysis. To expand the range of measures considered, it therefore will be necessary to restrict the sample to a subset of these studies, which in turn will reduce the number of degrees of freedom and the statistical power for the analyses. Yet another way to deal with this problem is to pool data for responses to survey questions that were asked in a similar, but not identical, way across studies.

In addition to exploring other measures of program characteristics and program impacts on other outcomes, it also will be possible to examine potential threshold effects for a small number of *specific* program characteristics and examine potential interactions for a small number of *combinations* of characteristics.

The simplest way to test for a threshold effect with respect to a program characteristic is to convert its continuous measure to a series of categorical indicator variables and substitute them into the program impact model. The estimated coefficients for these indicator variables can then be used to identify any sharp changes in program impacts that might occur when moving from one category to the next.[4]

The simplest way to test for interactions is to add interaction terms to the impact model. A two-way interaction term can be constructed as the product of measures for two program characteristics; a three-way interaction term can be constructed as the joint product of measures of three characteristics, and so on.[5] The estimated coefficient for each interaction term measures the magnitude and sign of its effect. This, in turn, indicates how the effect of one program characteristic on program impacts varies with the value of another characteristic.

Because testing for threshold effects and interactions requires *adding office-level variables* to the model, the number of such nonlinear effects that can be studied will be seriously limited by the fact that there are only 59 offices in the present sample. Consequently, it will be necessary to drastically reduce the number of potential candidate variables through a careful assessment of the *a priori* arguments for each. Doing so will help to mitigate the well-known pitfalls of data mining.

---

[4] In principle, this approach could be extended to include "spline functions" (Greene, 1997, pp. 387-390), but doing so is probably beyond the capacity of the current dataset.

[5] It is not likely that higher-order interactions will be useful or feasible given the limited number of program offices in the present sample and the difficulty of interpreting findings for such interactions.

The fourth major extension of the present analysis will be an attempt to estimate the *direct and indirect effects* of program characteristics on program impacts. However, this will be quite complicated because, in principle, it involves overlaying a system of simultaneous equations on a hierarchical model. Nevertheless, it should be possible to explore some of the implications of this issue by estimating a series of models that first examine the separate relationships among program impacts, program service differentials, and other program characteristics, and then piecing together the implications of these findings.[6]

In closing, it is important to note that the most difficult-to-overcome limitation of the present analysis is its potential exposure to left-out variable bias. As noted earlier, even though unbiased estimates of program impacts are obtained from a randomized experiment for each office, the relationships between these impacts and program characteristics are estimated non-experimentally. Hence, they are only as valid as the model upon which they are based. To the extent that the rich array of data available for the present analysis accounts for the most relevant determinants of program impacts, findings from the model provide valid causal inferences.[7] However, to the extent that such variables remain outside of the analysis, their influence will be attributed mistakenly to the variables that are included, thus biasing their causal inferences.

In theory, the ideal way to eliminate this problem is to randomly assign individual sample members to one of many different configurations of program characteristics. This ideal design (which would require a very large number of random assignment groups) would make it possible to distinguish the effectiveness of different program strategies for different target groups under different conditions. Although this is ideal in theory, it is not feasible in practice.

To date the random assignment approach has been used effectively to study a very small number of program alternatives (for example, standard versus reduced caseloads in GAIN; and integrated versus traditional case management strategies, and human capital development versus labor force attachment approaches in NEWWS). Thus it has demonstrated a clear potential for comparing a few specific alternatives. Nevertheless, the random assignment approach is not likely ever to be feasible for identifying and separating the influences of the many different forces, and combinations of forces, that impinge on human service programs.

Therefore, even with its limitations, the present approach of using natural cross-site variation in the characteristics of programs tested by randomized experiments ultimately may be the best *feasible* way to identify the factors that produce effective human service programs. To do this well, however, will require a large number of program sites, a conscious effort to conceptualize and measure relevant program characteristics consistently across these sites, and a well-developed modeling strategy for relating the variation in these characteristics to program impacts.

---

[6] This analysis could proceed by analogy to a fully recursive model with one independent variable, X, one intervening variable, Z, one dependent variable, Y, and independent error terms. See Baron and Kenny (1986) for a discussion of such an analysis.

[7] As noted earlier, relevant variables are ones that produce left-out variable bias because they are conditionally correlated with both program impacts and program characteristics in the model.

REFERENCES

Bane, Mary Jo. 1989. "Welfare Reform and Mandatory Versus Voluntary Work: Policy Issue or Management Problem?" *Journal of Policy Analysis and Management* 8(2): 285-89.

Barnow, Burt S. 1979. "Theoretical Issues in the Estimation of Production Functions in Manpower Programs." In *Evaluating Manpower Training Programs*, Research in Labor Economics, Supplement 1, (JAI Press): 295-338.

Baron, R. M., and David A. Kenny. 1986. "The Moderator-mediator Variable Distinction in Social Psychological Research." *Journal of Personality and Social Psychology* 51(6): 1173-1182.

Behn, Robert. 1991. *Leadership Counts: Lessons for Public Managers from the Massachusetts Welfare, Training and Employment Program.* Cambridge, MA: Harvard University Press.

Betsey, Charles, Robinson Hollister, and Mary Papageorgiou. 1985. *Youth Employment and Training Programs: The YEDPA Years.* Washington, DC: Committee on Youth Employment Programs, Commission on Behavioral and Social Sciences and Education, National Research Council, National Academy Press.

Bryk, Anthony, and Stephen Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.

Bryk, Anthony, Yeow Meng Thum, John Q. Easton, and Stuart Luppescu. 1998. "Academic Productivity of Chicago Public Elementary Schools." Chicago: Consortium on Chicago School Research.

Cook, Thomas, and Donald Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Chicago: Rand McNally.

Dehejia, Rajeev. 2000. "Was There a Riverside Miracle? A Framework for Evaluating Multi-Site Programs." New York: NBER Working paper #7844, August.

Economic Report of the President. 2000. Washington, DC: U.S. Government Printing Office.

Freedman, Stephen, Daniel Friedlander, Gayle Hamilton, JoAnn Rock, Marisa Mitchell, Jodi Nudelman, Amanda Schweder, and Laura Storto. 2000. *National Evaluation of Welfare-to-Work Strategies: Evaluating Alternative Welfare-to-Work Approaches: Two Year Impacts for Eleven Programs.* Washington DC: US Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation and US Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education, June.

Freedman, Stephen, Daniel Friedlander, Winston Lin, and Amanda Schweder. 1996. *The GAIN Evaluation: Five-Year Impacts on Employment, Earnings and AFDC Receipt*. New York: MDRC, July.

Friedlander, Daniel. 1988. *Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs.* New York: MDRC.

Friedlander, Daniel, and Gary Burtless. 1995. *Five Years After: The Long-Term Effects of Welfare-to-Work Programs.* New York: Russell Sage Foundation.

Friedlander, Daniel, David Greenberg, and Philip Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35: 1809-55.

Gladwell, Malcolm. 2000. *The Tipping Point.* Boston: Little, Brown and Company.

Greenberg, David, Robert Meyer, and Michael Wiseman. 1994. "Multi-site Employment and Training Evaluations: A Tale of Three Studies." *Industrial and Labor Relations Review* 47(4): 679-91.

Greenberg, David, and Mark Shroder. 1997. *Digest of Social Experiments*, 2nd edition. Washington, DC: The Urban Institute Press.

Greene, William H. 1997. *Econometric Analysis,* 3rd edition. Upper Saddle River, NJ: Prentice Hall.

Gueron, Judith M., and Edward Pauly. 1991. *From Welfare to Work*. New York: Russell Sage Foundation.

Hamilton, Gayle, and Thomas Brock. 1994. *The JOBS Evaluation: Early Lessons from Seven Sites.* Washington DC: US Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation and US Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education.

Hasenfeld, Yeheskel. 1983. *Human Service Organizations.* Englewood Cliffs, NJ: Prentice-Hall.

Hasenfeld, Yeheskel, and Richard English, eds. 1974. *Human Service Organizations: A Book of Readings*. Ann Arbor: University of Michigan Press.

Hasenfeld, Yeheskel, and Dale Weaver. 1996. "Enforcement, Compliance, and Disputes in Welfare-to-Work Programs." *Social Service Review* 70(2): 235-256.

Heckman, James J., Carolyn Heinrich, and Jeffrey Smith. 1999. "Understanding Incentives in Public Organizations." Presented at the National Academy of Sciences Colloquium, Devising Incentives to Promote Human Capital, December 17-18.

Heckman, James J., Jeffrey A. Smith, and Christopher Taber. 1996. "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program." In Gary D. Libecap, ed. *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth, Vol 7, Reinventing Government and the Problem of Bureaucracy*. Greenwich, CT: JAI Press, Inc.

Heinrich, Carolyn J. 2001. "Outcomes-based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness." Revision of paper presented at the annual meeting of the Association for Public Policy Analysis and Management, November 2000.

Heinrich, Carolyn J., and Laurence E. Lynn, Jr. 2000. "Governance and Performance: The Influence of Program Structure and Management on Job Training Partnership Act (JTPA) Program Outcomes." In Carolyn J. Heinrich and Laurence E. Lynn, Jr., eds. *Governance and Performance: New Perspectives*. Washington, DC: Georgetown University Press.

Hotz, V. Joseph, Guido W. Imbens, and Jacob A. Kerman. 2000. "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program." New York: NBER Working paper #W8007, November.

Kemple, James, Daniel Friedlander, and Veronica Fellerath. 1995. *Florida's Project Independence: Benefits, costs, and two-year impacts of Florida's JOBS program*. New York: Manpower Demonstration Research Corporation.

Kemple, James, and Joshua Haimson. 1994. *Florida's Project Independence: Program Implementation, Participation Patterns, and First-Year Impacts*. New York: Manpower Demonstration Research Corporation.

Kemple, James J., and Jason C. Snipes. 2000. *Career Academies: Impacts on Students' Engagement and Performance in High School*. New York: MDRC.

Kornfeld, Robert, and Howard S. Bloom. 1999. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics* 17(1): 168-197.

Lipsky, Michael. 1980. *Street-Level Bureaucracy*. New York: Russell Sage Foundation.

Lynn, Laurence E. Jr., Carolyn J. Heinrich, and Carolyn J. Hill. 2001. *Improving Governance: A New Logic for Empirical Research*. Washington, DC: Georgetown University Press.

Mead, Lawrence M. 1983. "Expectations and Welfare Work: WIN in New York City", *Policy Studies Review* 2(4): 648-661.

Mead, Lawrence M. 1986. *Beyond Entitlement: The Social Obligations of Citizenship*. New York: The Free Press.

Mead, Lawrence. 1989. "The Logic of Workfare: The Underclass and Work Policy." *Annals of the American Academy of Political and Social Science* 501: 156-170.

Meyer, Robert H. 1997. "Value-Added Indicators of School Performance: A Primer." *Economics of Education Review* 16(3): 283-301.

Michalopoulos, Charles, Christine Schwartz, with Diana Adams-Ciardullo. 2001. *National Evaluation of Welfare-to-Work Strategies: What Works Best for Whom? Impacts of 20 Welfare-to-Work Programs by Subgroup*. Washington DC: US Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation and US Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education, January.

Miller, Gary. 1992. *Managerial Dilemmas: The Political Economy of Hierarchy*. New York: Cambridge University Press.

Nathan, Richard. 1993. *Turning Promises Into Performance: The Management Challenge of Implementing Workfare*. New York: Columbia University Press.

Nunnally, Jum C. 1967. *Psychometric Theory*. New York: McGraw-Hill.

Pindyck, Robert S., and Daniel L. Rubinfeld. 1998. *Econometric Models and Economic Forecasts*, 4th edition. Boston: Irwin, McGraw-Hill.

Raudenbush, Stephen, Anthony Bryk, Yuk Fai Cheong, and Richard Congdon. 2000. *HLM5: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, Stephen W., and J. Douglas Willms. 1995. "The Estimation of School Effects." *Journal of Educational Behavioral Statistics* 20(4): 307-335.

Riccio, James, Howard S. Bloom, and Carolyn J. Hill. 2000. "Management, Organizational Characteristics, and Performance: The Case of Welfare-to-Work Programs." In Carolyn J. Heinrich and Laurence E. Lynn, Jr., eds. *Governance and Performance: New Perspectives*. Washington, DC: Georgetown University Press.

Riccio, James, and Daniel Friedlander. 1992. *GAIN: Program Strategies, Participation Patterns, and First-Year Impacts in Six Counties*. New York: Manpower Demonstration Research Corporation.

Riccio, James, Daniel Friedlander, and Stephen Freedman. 1994. *GAIN: Benefits, Costs, and Three-Year Impact of a Welfare-to-Work Program*. New York: MDRC.

Riccio, James, and Yeheskel Hasenfeld. 1996. "Enforcing a Participation Mandate in a Welfare-to-Work Program." *Social Service Review* 70(4): 516-42.

Riccio, James, and Alan Orenstein. 1996. "Understanding Best Practices for Operating Welfare-to-Work Programs." *Evaluation Review* 20(1): 3-28.

Sanders, William L., and Sandra P. Horn. 1994. "The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment." *Journal of Personnel Evaluation in Education* 8: 299-311.

Sherwood, Kay, and Fred Doolittle. 1999. "What's Behind the Impacts: Doing Implementation Research in the Context of Program Impact Studies." Working paper: New York: MDRC.

Smith, Jeffrey and Miana Plesca. Forthcoming. "How Can We Improve Public Employment and Training Programs? Working paper, University of Western Ontario.

Table 1

GAIN, PI, and NEWWS[a]

	GAIN: Greater Avenues for Independence	PI: Project Independence	NEWWS: National Evaluation of Welfare-to-Work Strategies
Period of Random Assignment	March 1998 to June 1990	January 1991 to August 1991[b]	June 1991 to December 1994
Number of Offices	22	10	27
Number of Counties	6	9	10
States	California	Florida	California, Georgia, Michigan, Ohio, Oklahoma, Oregon

Notes:

a. The sample in this analysis is restricted to females only.

b. PI random assignment occurred from July 1990 to August 1991. However, the present study restricts the PI analysis sample to only those clients randomly assigned in 1991 because program conditions changed in 1991, and the staff survey used to measure program characteristics was conducted in September and October 1991 (see Kemple, Friedlander, and Fellerath, 1995; Kemple and Haimson, 1994). Therefore, the office-level measures most accurately reflect conditions faced only by clients randomly assigned in 1991.

Table 2

Sample Sizes

	GAIN	PI	NEWWS	TOTAL
Total number of program offices	22	10	27	59
Total experimental sample	18,126	4,296	46,977	69,399
Total program staff survey sample	776	57	392	1,225
Total client follow-up survey sample	3,163	692	11,380	15,235
Experimental sample per office				
Mean	824	430	1,740	1,176
Standard deviation	488	166	1,390	1,117
Range	260 to 2,212	177 to 764	289 to 4,418	177 to 4,418
Staff survey sample per office				
Mean	35	6	15	21
Standard deviation	19	2	14	19
Range	9 to 83	2 to 8	1 to 61	1 to 83
Client follow-up survey sample per office[a]				
Mean	144	69	421	258
Standard deviation	143	11	574	423
Range	36 to 656	54 to 91	27 to 2,159	27 to 2,159

Note:

[a] The client follow-up survey was used to measure employment and training services received by program and control group members.

Table 3
Estimated Program Impacts on Mean Total Earnings During the First Two Years After Random Assignment By Local Program Office

Office	Impact Estimate[a]	Standard Error Of the Estimate[a]	Statistical Significance of the Estimate (p-value)
GAIN1	$ 4,217	$ 1,023	0.000
NEWWS1	3,775	1,212	0.002
GAIN2	2,968	655	0.000
GAIN3	2,904	1,201	0.016
GAIN4	2,765	480	0.000
NEWWS2	2,679	1,157	0.021
NEWWS3	2,486	628	0.000
GAIN5	2,261	789	0.004
NEWWS4	1,914	549	0.001
GAIN6	1,779	983	0.070
NEWWS5	1,758	333	0.000
GAIN7	1,740	803	0.030
GAIN8	1,681	765	0.028
PI1	1,668	1,062	0.116
NEWWS6	1,596	391	0.000
NEWWS7	1,573	792	0.047
NEWWS8	1,422	1,068	0.183
NEWWS9	1,404	563	0.013
GAIN9	1,376	1,115	0.217
NEWWS10	1,369	566	0.016
PI2	1,219	897	0.174
GAIN10	1,182	729	0.105
GAIN11	1,143	532	0.032
GAIN12	1,113	811	0.170
NEWWS11	1,049	1,074	0.329
NEWWS12	925	295	0.002
NEWWS13	899	760	0.237
PI3	894	1,194	0.454
NEWWS14	811	292	0.006
NEWWS15	759	485	0.118
GAIN13	700	1,006	0.487
NEWWS16	609	365	0.096
NEWWS17	573	271	0.035
GAIN14	556	930	0.550
NEWWS18	531	273	0.052

(continued)

Table 3
Estimated Program Impacts on Mean Total Earnings During the First Two Years After Random Assignment By Local Program Office
(continued)

Office	Impact Estimate[a]	Standard Error Of the Estimate[a]	Statistical Significance of the Estimate (p-value)
PI4	525	1,614	0.745
NEWWS19	494	294	0.093
NEWWS20	481	844	0.569
PI5	437	1,112	0.694
GAIN15	371	523	0.478
PI6	310	838	0.711
NEWWS21	309	412	0.454
PI7	306	1,028	0.766
NEWWS22	200	327	0.540
NEWWS23	175	558	0.754
GAIN16	100	1,215	0.934
NEWWS24	-140	473	0.768
PI8	-201	1,057	0.849
GAIN17	-226	635	0.722
NEWWS25	-338	405	0.404
GAIN18	-342	670	0.610
GAIN19	-356	1,281	0.781
PI9	-372	956	0.697
PI10	-716	732	0.328
GAIN20	-754	845	0.373
NEWWS26	-884	725	0.223
NEWWS27	-942	856	0.271
GAIN21	-1,233	994	0.215
GAIN22	-1,412	951	0.137
Average	883		
Standard Deviation	1,182		
Range	-1,412 to 4,217		

Note:
[a] Values are in 1996 dollars.

Table 4:

Survey Items for the Management Scales Related to Service Technology

Scale and Items[a]

Emphasis on moving clients into jobs quickly

- *Does* your unit emphasize helping clients build basic skills, or moving them quickly into jobs?
- *Should* your unit emphasize helping clients build basic skills, or moving them quickly into jobs?
- What would be your *personal advice* to a client who can either take a low-skill, low-paying job OR stay on welfare and wait for a better opportunity?
- What advice would *your supervisor* want you to give to such a client?

Emphasis on personalized client attention

- Does your program emphasize the quality of its services more than the number of clients it serves?
- During intake, does your unit spend enough time with clients?
- During intake, do staff make an effort to learn about clients' family problems?
- During intake, do staff make an effort to learn about clients' goals and motivation to work?
- How well is your program tailoring services to clients' needs?

Closeness of client monitoring

- How closely are staff monitoring clients?
- If a client has been assigned to adult basic education but has not attended, how soon would staff find out?
- If a client has been assigned to vocational education but has not attended, how soon would staff find out?
- How closely is your agency monitoring whether clients quit or lose part-time jobs?
- Once your agency learns a client lost a part-time job, how soon would she be assigned to another activity?

Note:

a. The questions in this table paraphrase each staff survey question. See Appendix Table C3 for the exact wording of each question and its response scale.

Table 5

Summary of Local Program Characteristics

Program Characteristic	Mean	Standard Deviation	Range
Program Management			
Emphasis on moving clients into jobs quickly	0.0	1.0	-1.7 to 2.5
Emphasis on personalized client attention	0.0	1.0	-2.0 to 2.3
Closeness of client monitoring	0.0	1.0	-2.8 to 1.9
Staff caseload size	136	67	70 to 367
Frontline staff/supervisor inconsistency about service technology	0.0	1.0	-1.5 to 3.2
Frontline staff inconsistency about service technology	0.0	1.0	-2.1 to 4.5
Service Differential			
Basic education	11	13	-11 to 50
Job search assistance	17	12	-13 to 47
Vocational training	5	10	-21 to 35
Economic Environment			
Average monthly unemployment rate (in percent)	7.4	3.1	3.5 to 14.3

Table 6

Summary of Service Receipt Rates and Service Differentials

	Basic Education	Job Search Assistance	Vocational Training
Mean Percentage of Program Group Members Who Received Service	19	22	27
Mean Percentage of Control Group Members Who Received Service	8	5	22
Mean *Service Differential* (the Program/ Control Group *Difference* in Service Receipt Rates)	11	17	5
Standard Deviation Across Offices of the Service Differential	13	12	10
Range Across Offices of the Service Differential	-11 to 50	-13 to 47	-21 to 35

BEST COPY AVAILABLE

Table 7

Client Characteristics[a]

At Random Assignment the Sample Member:	Percent of Full Sample of Individuals	Cross-Office Standard Deviation	Cross-Office Range (Percent)
Was a high school graduate or had a GED	56	14	17 to 74
Had one child	42	6	30 to 56
Had two children	33	3	28 to 50
Had three or more children	25	6	11 to 39
Had a child under six years old	46	23	7 to 73
Was less than 25 years old	19	11	1 to 42
Was 25 to 34	49	7	23 to 57
Was 35 to 44	26	8	14 to 45
Was 45 or older	6	6	2 to 34
Was White, non-Hispanic	41	24	1 to 87
Was Black, non-Hispanic	41	27	0 to 98
Was Hispanic	14	22	0 to 92
Was Native American	2	3	0 to 21
Was Asian	2	3	0 to 23
Was some other race/ethnicity	<1	1	0 to 5
Was a welfare applicant	17	31	0 to 99
Had received welfare continuously for the past 12 months	44	27	0 to 96
Had no earnings in the past year	56	13	29 to 81
Had earned $1 to $2499	21	5	10 to 30
Had earned $2500 to $7499	14	5	6 to 26
Had earned $7500 or more	9	6	2 to 27
Sample size	69,399		

Note:

a. The sample in this analysis is restricted to females only.

Table 8

Effects of Local Program Characteristics On Program Impacts[a]

Local Program Characteristic	Regression Coefficient[b]	Partially Standardized Regression Coefficient[b]	Statistical Significance of Coefficient (p-value)	Conditional Impact Interval (in Dollars)[b]	Conditional Impact Interval (in Percent)
Program Management					
Emphasis on moving clients into jobs quickly	$ 720	$ 720	0.000	$ 397 to 1,361	8.1 to 27.9
Emphasis on personalized client attention	428	428	0.000	592 to 1,166	12.2 to 23.9
Closeness of client monitoring	-197	-197	0.110	1,011 to 747	20.8 to 15.3
Staff caseload size	-4	-268	0.003	1,058 to 700	21.7 to 14.4
Frontline staff/ supervisor inconsistency about service technology	-159	-159	0.102	986 to 772	20.2 to 15.9
Frontline staff inconsistency about service technology	124	124	0.141	796 to 962	16.3 to 19.8
Service Differential					
Basic education	-16	-205	0.017	1,017 to 741	20.9 to 15.2
Job search assistance	1	12	0.899	871 to 887	17.9 to 18.2
Vocational training	7	71	0.503	831 to 927	17.1 to 19.0
Economic Environment					
Unemployment rate	-94	-291	0.004	1,074 to 684	22.1 to 14.0
Mean Program Impact on Earnings	879		0.000		
Mean Counterfactual	4,871		0.000		
Impact as Percent of Counterfactual	18.0				

Notes:

a. Results in this table are based on a sample of 69,399 program and control group members from 59 local welfare-to-work program offices.

b. Values are in 1996 dollars.

Table 9

Relationships Between Client Characteristics and Program Impacts[a]

At Random Assignment the Sample Member:	Regression Coefficient[b]	Statistical Significance (p-value)
Was a high school graduate or had a GED	$ 653	0.001
Had one or no children (left-out)		
Had two children	301	0.160
Had three or more children	591	0.003
Had a child under six years old	34	0.841
Was less than 25 years old	206	0.557
Was 25 to 34	105	0.707
Was 35 to 44	305	0.376
Was 45 or older (left out)		
Was White, non-Hispanic (left-out)		
Was Black, non-Hispanic	-178	0.369
Was Hispanic	-213	0.527
Was Native American	-696	0.115
Was Asian	353	0.560
Was some other race/ethnicity	726	0.487
Was a welfare applicant	-145	0.532
Had received welfare continuously for the past 12 months	444	0.085
Had zero earnings in the past year (left-out)		
Had earned $1 to $2499	-186	0.222
Had earned $2500 to $7499	72	0.787
Had earned $7500 or more	22	0.965
Mean Program Impact on Earnings	879	0.000
Mean Counterfactual	4,871	0.000
Impact as Percent of Counterfactual	18.0	

Notes:

a: Results in this table are based on a sample of 69,399 program and control group members from 59 local welfare-to-work program offices.

b. Coefficient estimates are in 1996 dollars.

Table 10

Program Impacts on Other Outcomes to be Examined in Future Research[a]

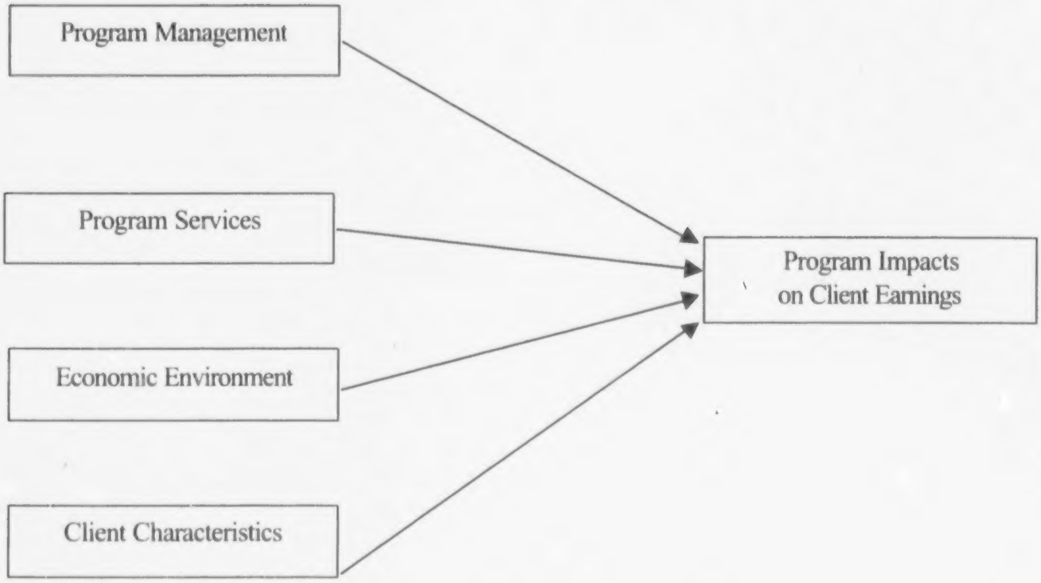
Impact	Mean Impact	Mean Percentage Impact	Statistical Significance of Mean Impact (p-value)	Range of Unconditional Impact Estimates Across Offices	Statistical Significance of the Variation in Unconditional Impacts Across Offices (p-value)	Statistical Significance of the Variation in Conditional Impacts Across Offices (p-value)
Total Impacts for the Two-Year Follow-up Period						
Number of Quarters Employed	0.36	15.7	0.000	-0.19 to 0.92	0.000	0.000
Total AFDC Payments (in 1996 dollars)	-757	8.0	0.000	-1,712 to 353	0.000	0.000
Number of Quarters Receiving AFDC	-0.30	5.1	0.000	-0.61 to -0.04	0.000	0.000
Impacts During the Eighth Quarter After Random Assignment						
Earnings (in 1996 dollars)	120	15.7	0.000	-25 to 332	0.000	0.013
Percent Employed	4.6	14.6	0.000	-2.3 to 10.8	0.000	0.000
AFDC Payments (in 1996 dollars)	-99	10.6	0.000	-191 to -23	0.000	0.000
Percent Receiving AFDC	-4.3	13.7	0.000	-10.4 to -0.6	0.000	0.001

Note:

a. Results in this table are based on a sample of 69,399 program and control group members from 59 local welfare-to-work program offices.

Figure 1

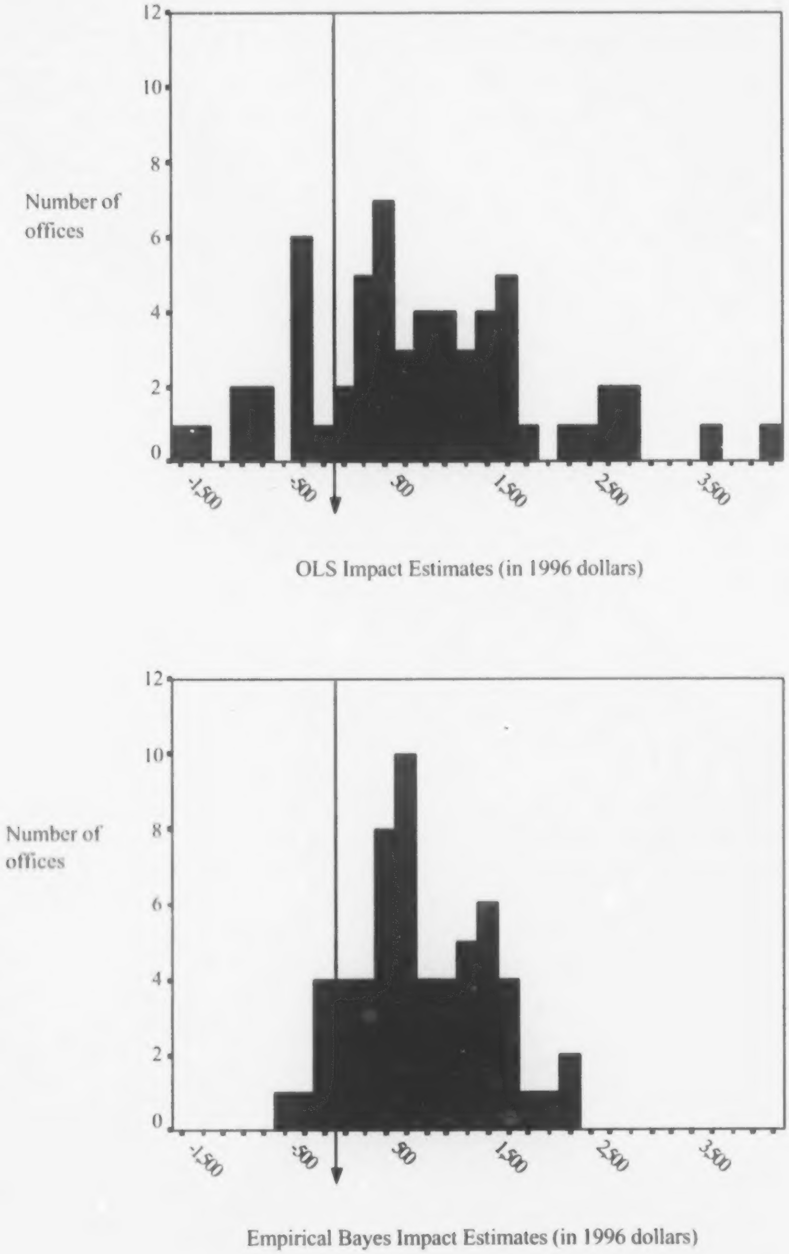
How Welfare-to-Work Programs Affect Client Earnings[a]


Program Management
Program Services
Economic Environment
Client Characteristics
Program Impacts on Client Earnings

Note:

a. This depiction is greatly simplified. Undoubtedly, complex interdependencies exist among these factors.

Figure 2
The Distributions of Unconditional OLS Impact Estimates and Empirical Bayes Impact Estimates



Appendix A

The Program Models for GAIN, PI, and NEWWS

GAIN served as California's JOBS program. In contrast to earlier welfare-to-work initiatives, GAIN was noted particularly for the importance it placed on basic education for those determined to need remediation in basic reading or math skills or instruction in English as a Second Language. The program also provided job search assistance, unpaid work experience, and referrals for post-secondary education and vocational training (Riccio and Friedlander, 1992). Participation in GAIN, as in all JOBS programs, was mandatory for a large part of the welfare caseload.[1] Those who were required to participate but failed to do so without what was considered to be "good cause" were to receive a financial sanction, i.e., a penalty in the form of a reduction in the family's welfare grant. Welfare recipients who were mandated to participate in GAIN and who attended a GAIN orientation were randomly assigned to either a program group or a control group; random assignment occurred at the GAIN orientation session. Appendix Figure A1 illustrates the flow of steps in the GAIN program model.[2]

PI served as Florida's JOBS program. Although it, like GAIN, provided a range of activities and services and included a similar participation mandate, it focused particularly on low-cost job search strategies and more limited access to basic education (Kemple and Haimson, 1994). Also in contrast to GAIN, random assignment occurred at an earlier point in PI, when individuals first applied for AFDC benefits or when their benefit eligibility was determined or redetermined. Appendix Figure A2 illustrates the flows of steps in the PI program model.

Sharp cutbacks in the provision of PI services during the second half of the original study period are a distinguishing aspect of PI's implementation, which has important implications for the present analysis. These cutbacks were a response to funding reductions, a staff hiring freeze, and rapid welfare caseload growth that strained limited existing resources (Kemple, Friedlander, and Fellerath, 1995, pp. 6-9). Thus, two enrollment cohorts, which were exposed to two essentially different PI programs, were identified by the original evaluation. The "early" cohort contained sample members who were enrolled and randomly assigned between July and December 1990; the "late" cohort contained sample members who were enrolled and randomly assigned between January and August 1991. Only the late cohort was included in the present analysis because the staff survey used to measure program characteristics was administered only during the time when this group was in the program (see Appendix C).

---

[1] Mandatory GAIN recipients were originally defined as "single parents whose youngest child was six or older and the heads of two-parent households" (Riccio and Friedlander, 1992, p. 11). The single parents in this group make up most of the GAIN sample members included in the present study. However, it should be noted that when GAIN became California's JOBS program, the mandatory population was expanded to include those whose youngest child was at least 3 years of age.

[2] As noted later, Riverside GAIN offices conducted a special caseload size experiment to study the effect of varying the caseload size. The present analysis does not include the sample members from this study.

NEWWS is a six-state evaluation of alternative welfare-to-work strategies. The programs included in this evaluation offered a range of activities similar to those offered by GAIN and PI. All NEWWS sites have operated under JOBS program rules, and, in all but two sites, random assignment was conducted at the point of orientation for the JOBS program. In two sites, however, random assignment was conducted at the point when clients applied for AFDC benefits or had their eligibility redetermined (as in PI).

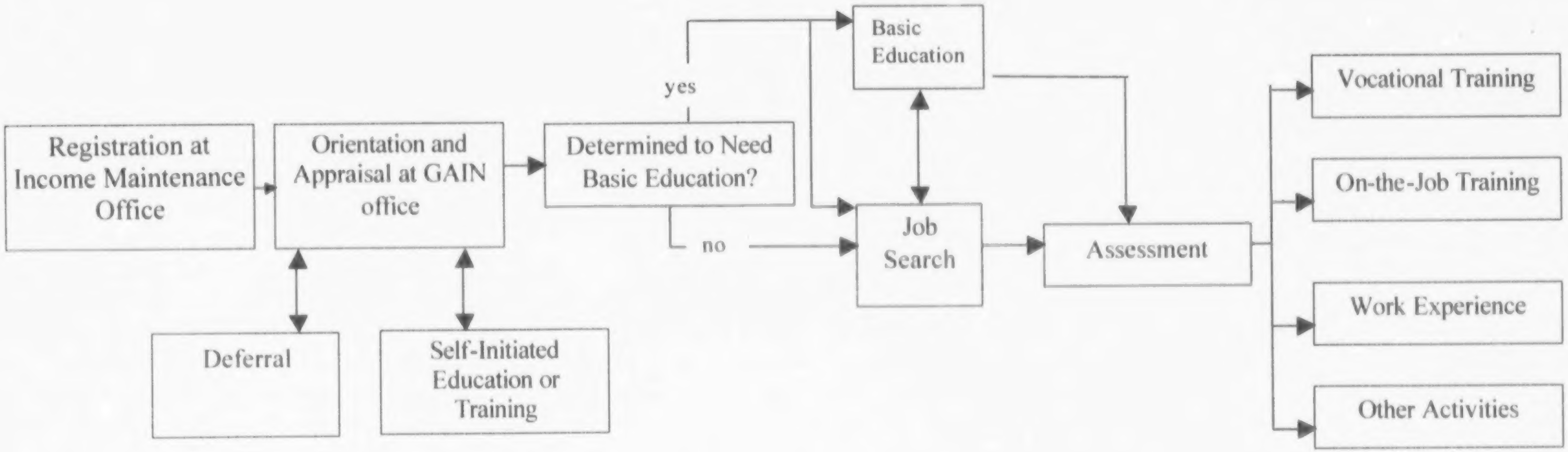
Perhaps the most unique feature of NEWWS is that three of its sites—Atlanta, Georgia; Grand Rapids, Michigan; and Riverside, California—implemented a three-way random assignment design in order to permit direct comparisons of the effectiveness of labor force attachment and human capital development strategies relative to a common control group (see Table A1).

A fourth site, Columbus, Ohio, implemented a three-way random assignment design to assess different ways of organizing case management. Clients at this site were randomly assigned either to one program group for which case management and income maintenance tasks were performed by separate case managers (the "traditional" strategy); a second program group for which case management and income maintenance tasks were performed by the same case manager (an "integrated" strategy); or a control group, which was not subject to JOBS program requirements and did not receive JOBS services. Both program groups placed a substantial emphasis on basic education and job skills training.

The three other NEWWS sites—Detroit, Michigan; Oklahoma City Oklahoma; and Portland, Oregon—randomly assigned clients to their JOBS program or a control group. The Detroit and Oklahoma City programs were mainly education-focused, while the Portland program emphasized labor force attachment with a mix of education and training activities.

Figure A1

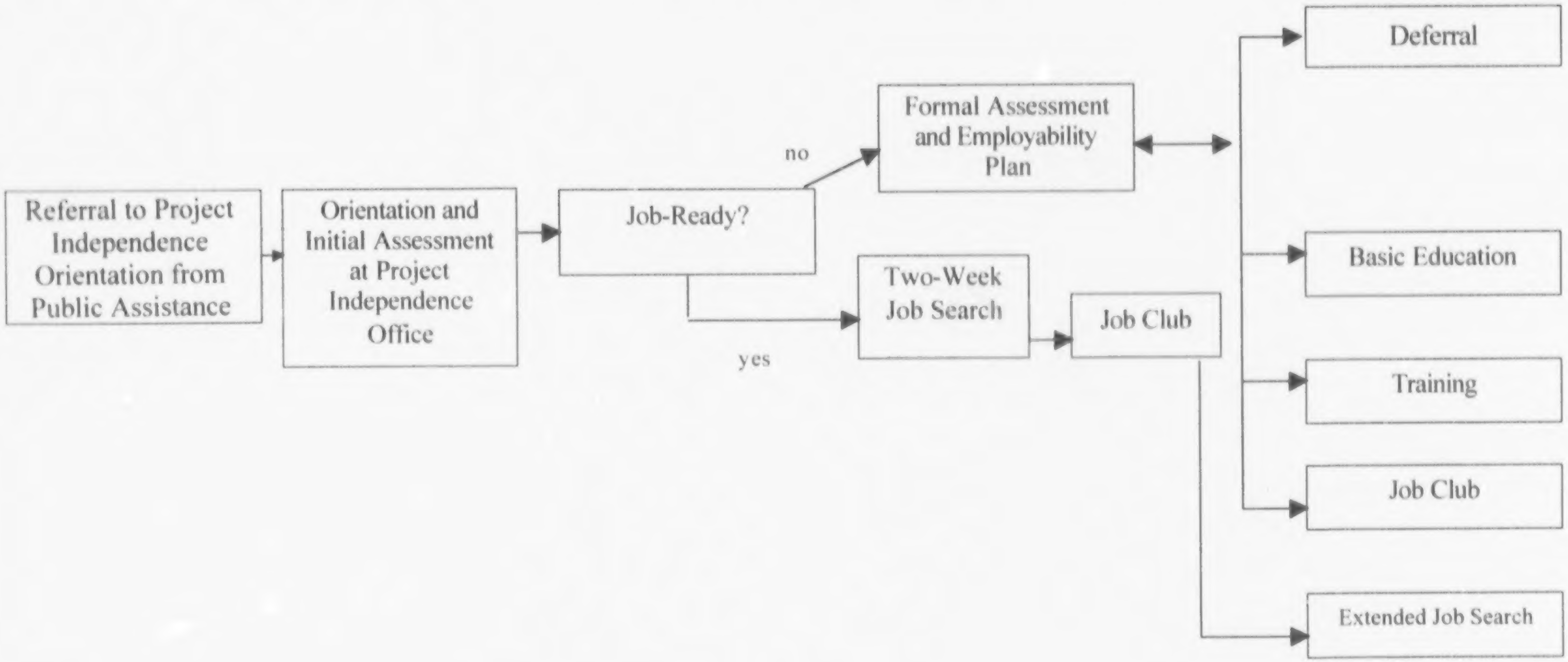
The GAIN Program Model


Registration at Income Maintenance Office
Orientation and Appraisal at GAIN office
Determined to Need Basic Education?
yes
no
Basic Education
Job Search
Assessment
Deferral
Self-Initiated Education or Training
Vocational Training
On-the-Job Training
Work Experience
Other Activities

SOURCE: Adapted from Riccio and Friedlander (1992), p. 4.

Figure A2

The PI Program Model


Referral to Project Independence Orientation from Public Assistance
Orientation and Initial Assessment at Project Independence Office
Job-Ready?
no
yes
Formal Assessment and Employability Plan
Two-Week Job Search
Job Club
Deferral
Basic Education
Training
Job Club
Extended Job Search

SOURCE: Adapted from Kemple and Haimson (1994), p. xviii.

Table A1

The NEWWS Program Models

Characteristic	Atlanta, GA	Grand Rapids, MI	Riverside, CA	Columbus, OH	Detroit, MI	Oklahoma City, OK	Portland, OR
Type of random assignment	Three-way (2 program groups, 1 control group)	Three-way (2 program groups, 1 control group)	Three-way (2 program groups, 1 control group)	Three-way (2 program groups, 1 control group)	Two-way (1 program group, 1 control group)	Two-way (1 program group, 1 control group)	Two-way (1 program group, 1 control group)
Point of random assignment	Program orientation	Program orientation	Program orientation	Welfare application or redetermination	Program orientation	Welfare application	Program orientation
Type of study	Differential impacts of LFA and HCD approaches	Differential impacts of LFA and HCD approaches	Differential impacts of LFA and HCD approaches	Differential impacts of integrated and traditional case management strategies	Net impacts of established program	Net impacts of established program	Net impacts of established program
Employment-focused approach	Yes: LFA group	Yes: LFA group	Yes: LFA group				Yes
Education-focused approach	Yes: HCD group	Yes: HCD group	Yes: HCD group	Yes: both integrated and traditional groups	Yes	Yes	

Source: adapted from Exhibits ES-1 and ES-2 (Freedman et al., 2000).

Notation: LFA: Labor Force Attachment
HCD: Human Capital Development

Appendix B

Measuring Program Performance as Program Impacts on Client Earnings

This appendix describes how the performance of welfare-to-work program offices was measured, where "performance" is defined as impact on mean client earnings for the first two years after random assignment. The appendix also examines the statistical significance of the variation in impacts across program offices, and the extent to which this variation is "explained" by client characteristics and program characteristics.

B.1 Linking Program and Control Group Members to Local Program Offices

An initial step needed to measure program impacts on client earnings is to link the program and control group members of the sample with the relevant local program office. Matching clients to offices is straightforward for clients in GAIN, in PI, and for most offices in NEWWS: program and control group members are linked to the local JOBS program office where program group members received services (and where control group members *would have received services* had they not been in the control group).

For the present analysis, a special procedure was needed to assign program and control group members to offices in two of the three NEWWS sites that conducted three-way random assignment (see Appendix A for descriptions of sites).[1] First, because different staff served the two program groups in these sites, the present analysis classified each of the two different program streams as a separate "local program office." For example, program group members in the Atlanta LFA stream were assigned to a unique office, and program group members in the Atlanta HCD stream were assigned to a different unique office. Second, the three-way random assignment procedure in a site created a single control group. To provide each of the local offices in these sites with an appropriate comparison group from which a counterfactual could be estimated in the regression analysis, the same control group members were used to construct impacts for both the LFA and HCD programs.[2] In the example above, the control group members in Atlanta were used once to estimate a counterfactual for the Atlanta LFA local program office, and again to estimate a counterfactual for the Atlanta HCD local program office.

B.2 Measuring Earnings

Earnings data for the analysis were obtained from state UI wage records, which report quarterly information on earnings from all jobs "covered" by unemployment insurance.[3] Earnings for each follow-up quarter (including zeros for quarters with no reported UI-covered earnings) were used to compute total earnings for the first two years after random assignment for each sample member.

---

[1] The exception is Grand Rapids. Because local office staff served both LFA and HCD program group members, only one Grand Rapids office (combining both LFA and HCD stream clients) was constructed.

[2] 5,266 control group members were used twice as if they were independent individuals, and no adjustments were made to the standard errors. Because this group constituted a small fraction of the full sample, not adjusting standard errors for their duplicate use likely had a negligible effect on the findings.

The random assignment dates for sample members vary both within and across offices, and two strategies were used to account for these timing differences. First, to assure that all earnings amounts are comparable over time, they were converted to constant 1996 dollars using the CPI-U (Economic Report of the President, 2000). Second, to align total earnings temporally the same way for all sample members, the measure of earnings used for each client was the sum of her earnings for the first eight quarters after her quarter of random assignment. Quarterly earnings were top-coded at $20,000 to reduce the potential for distortion produced by large data errors.

B.3 Estimating Program Impacts

Using comparable measures of total two-year follow-up earnings, it was possible to estimate the impact of the program at each local office as a regression-adjusted difference in the mean earnings of its program and control group members.

B.3.1 The Starting Point: Unconditional Program Impacts

Unconditional program impacts, which do not control for the relationships between client characteristics and program impacts, are the starting point for the present analysis. These estimates were obtained from the following model, which controls for chance differences between observable characteristics of control group and program group members:

$$Y_{ji} = \sum_{J} \alpha_J LO(J)_{ji} + \sum_{J} \tilde{\beta}_J LO(J)_{ji} P_{ji} + \sum_{k} \delta_k CC_{kji} + \sum_{J} \kappa_J LO(J)_{ji} RA_{ji} + \varepsilon_{ji} \qquad \textbf{(B1)}$$

where client characteristics are grand-mean centered(measured as the deviation from the mean value for the full sample), and:

Y_{ji} = total two-year follow-up earnings for sample member i from office j,
$LO(J)_{ji}$ = 1 if sample member i is from local office J and zero otherwise (with a separate indicator variable for each office J),
P_{ji} = one if sample member i from office j is a program group member and zero otherwise,
CC_{kji} = client characteristic k for sample member i from office j,
RA_{ji} = a zero/one indicator variable to distinguish members of two sample cohorts at office j that were subject to different random assignment ratios,

---

[3] Kornfeld and Bloom (1999) describe the types of data collected by the UI system and assess its validity for measuring earnings for low-income persons.

α_j = mean two-year follow-up earnings at office j for *the typical control group member from the full study sample,*

$\tilde{\beta}_j$ = the unconditional program impact at office j for the *typical program group member from that office,*

δ_k = a regression coefficient indicating how mean two-year follow-up earnings vary with client characteristic k,

κ_j = the regression-adjusted difference in mean follow-up earnings for control group members in the two random assignment cohorts at office j,

ε_{ji} = a random error term for sample member i from local office j, which is assumed to be independently and identically distributed across individual sample members.

Of primary interest in this model are estimates of the $\tilde{\beta}_j$: the program impact for the *typical sample member from each office*. Table 3 in the paper lists these unconditional impact estimates by office.[4] They range from a low of - $1,412 to a high of $4,217, with a mean of $883 and a standard deviation of $1,182 (all in 1996 dollars).[5] As discussed below in Section B.4, the variation in unconditional impacts across offices was highly statistically significant. Hence, there was real variation in impacts to be explained by differences in client characteristics and program characteristics.

B.3.2 The Next Step: Conditional Program Impacts

Conditional program impacts, which control for office differences in client characteristics, represent the next conceptual step in the analysis. These impacts can be estimated from the following model, which adds interactions between program/control status and client characteristics to Equation B1:

$$Y_{ji} = \sum_J \alpha_j LO(J)_{ji} + \sum_J \beta_j LO(J)_{ji} P_{ji} + \sum_k \delta_k CC_{kji} + \sum_k \gamma_k CC_{kji} P_{ji} + \sum_J \kappa_j LO(J)_{ji} RA_{ji} + \varepsilon_{ji} \qquad \textit{(B2)}$$

where the client characteristics are grand-mean centered and the variables are defined as in Equation B1, with the following exceptions:

β_j = the conditional program impact at office j for the *typical program group member from the full study sample,*

γ_k = a regression coefficient indicating how *impacts vary with client characteristic k.*

Because client characteristics are grand-mean centered, and the interaction terms in Equation B2 account for how program impacts vary with client characteristics, β_j represents the average impact for the *typical member of the full study sample* if she had been in the program at local office j. Thus, β_j is *conditioned* on client characteristics.[6] As explained below, the variation across offices in conditional impacts was also highly statistically significant. Hence, there was real variation to be explained by differences in program characteristics.

---

[4] These estimates were obtained using ordinary least squares.

[5] As stated in section 3.3, this mean of $883 weights each site equally and is presented for descriptive purposes only. It differs from the estimated grand mean impact, reported elsewhere in this paper, which weights each site according to the reliability of its impact estimate.

B.3.3 The "Complete" Impact Model: Including Program Characteristics

As described in Section 2.2 of the paper, the complete model of program impacts has a Level One component, plus a Level Two component consisting of three equations that describe how α_j, β_j and κ_j vary across offices. This model is presented below in notation for a hierarchical linear model, which is somewhat different than the notation for fixed effects models used in equations B1 and B2 above. The Level One and Level Two components are estimated simultaneously:

LEVEL ONE

$$Y_{ji} = \alpha_j + \beta_j P_{ji} + \sum_k \delta_k CC_{kji} + \sum_k \gamma_k CC_{kji} P_{ji} + \kappa_j RA_{ji} + \varepsilon_{ji} \qquad \textit{(B3)}$$

where all variables are defined as in Equation B2, client characteristics are grand-mean centered, and α_j, β_j, and κ_j vary randomly across offices.

LEVEL TWO

Conditional Program Impacts by Office

$$\beta_j = \beta_0 + \sum_m \pi_m PM_{mj} + \sum_n \phi_n PS_{nj} + \psi EE_j + \mu_j \qquad \textit{(B4)}$$

where all independent variables are grand-mean centered and:

β_j = the conditional program impact at local office j for the typical program group member from the full study sample,
PM_{mj} = program management variable m for local office j,
PS_{nj} = program service variable n for local office j,
EE_j = the economic environment variable for local office j,
β_0 = the grand mean program impact for a typical program group member from the full study sample,
π_m = the effect of program management feature m on program impacts,
ϕ_n = the effect of program service n on program impacts,
ψ = the effect of the economic environment on program impacts,
μ_j = a random component of the program impact for office j.

---

[6] To simplify the analysis and keep its computations manageable, the present model specifies interactions between client characteristics and program impacts that are constant across local offices. Future analyses will consider more complex specifications, but they are not likely to change the present findings appreciably.

Conditional Control Group Mean Earnings by Office

$$\alpha_j = \alpha_0 + \lambda EE_j + \upsilon_j \tag{B5}$$

where the economic environment variable is grand-mean centered and:

α_j = the conditional control group mean earnings at local office j for a typical control group member from the full study sample,
EE_j = the value of the economic environment variable for local office j,
α_0 = the grand mean conditional control group earnings for a typical member of the full study sample,
λ = the effect of the economic environment on control group earnings,
υ_j = a random component of the conditional control group mean earnings for office j.

Random Assignment Cohort Mean Earnings Differences by Office[7]

$$\kappa_j = \kappa_0 + \eta_j \tag{B6}$$

where:

κ_j = the difference in conditional mean earnings for the two random assignment cohorts at each office j (these cohorts were defined to account for changes that occurred over time in the random assignment ratios in local office j),
κ_0 = the grand mean difference in conditional mean earnings for the two random assignment cohorts, and
η_j = a random component of the difference in conditional mean earnings for the two random assignment cohorts at office j.

The complete model is discussed in Section 2.2 of the paper and compared to the conditional and unconditional models in Table B1.

B.4 Assessing the Statistical Significance of the Impact Variation

Statistically significant (real) variation in impacts across offices is a prerequisite for estimating the relationships between client characteristics, program characteristics and program impacts. Without real variation in impacts, there is no way to observe how impacts covary with client characteristics and program impacts. For example, if impact estimates are statistically significant and large for every office, but they are the same everywhere, there is no variation for client characteristics or program characteristics to explain, and thus no information to use to measure their influence on program impacts.

---

[7] See Section 2.2.3 for additional discussion of this component of the model.

Table B2 indicates that the variation in program impacts across offices in the present sample is, in fact, highly statistically significant. The first column lists the Chi-Square statistic used to test this hypothesis and the second column lists its corresponding significance level or p-value (Bryk and Raudenbush 1992, pp. 54-56).

The first row in the table indicates that unconditional program impacts vary significantly across offices; so there is real variation to be explained by client characteristics and program characteristics. The second row indicates that the conditional program impacts — that is, those obtained after controlling statistically for observable client characteristics — also vary significantly across offices, so there is still a statistically significant amount of variation in conditional impacts that could potentially be explained by program characteristics. The final row in the table indicates that even after controlling for both client characteristics and program characteristics, there is a statistically significant amount of variation that could be explained by factors that are not in the model.

B.5 "Explaining" the Impact Variation

At each stage of the preceding analysis there was a smaller and smaller amount of variation in impacts across offices to be explained by adding further independent variables to the model. One way to represent this phenomenon is through a quasi-"R-squared" statistic. For example, results of the analysis indicate that the variation in conditional impacts was 15.6 percent smaller than the variation in unconditional impacts.[8] Thus, client characteristics (the basis for defining conditional impacts) by themselves explain 15.6 percent of the variation in program impacts across offices. After adding program characteristics to the model, the remaining unexplained variation is 79.9 percent less than the original unconditional variation. Thus, client characteristics (in Level One of the model) *plus* program characteristics (in Level Two of the model) jointly explain 79.9 percent of the original unconditional impact variation. Hence, the complete two-level model has substantial explanatory power.

---

[8] Maximum Likelihood methods were used to estimate "real" variation in impacts, which is consistent with standard practice for hierarchical modeling (Bryk and Raudenbush, 1992, pp. 39-44).

Table B1

Independent Variables in the Unconditional, Conditional, and "Complete" Models of Program Impacts

	Model		
	Unconditional	**Conditional**	**Complete**
Independent Variables in Level One			
Client Characteristics	✓	✓	✓
Interactions Between Client Characteristics and Program/Control Status		✓	✓
Independent Variables in Level Two			
Economic Environment as a Predictor of Control Group Earnings (the Counterfactual)			✓
Program Management, Program Services, and Economic Environment as Predictors of Program Impacts			✓

Note: ✓ indicates that the type of variable is included in the model.

Table B2

Statistical Significance of the Variation in Program Impacts Across Offices

Impact Variation	Chi-Square Test Statistic	Statistical Significance (p-value)
Unconditional Impacts	146.16	0.000
Conditional Impacts	130.61	0.000
Residual Impacts from the Complete Model	61.63	0.089

Appendix C

Measuring Program Characteristics

This appendix describes how measures were constructed for the three types of program characteristics that are included in the present impact model: program management, program services and the local economic environment.

C.1 Constructing the Program Management Measures

First consider how the program management measures were constructed.

C.1.1 The Data Source: Local Staff Surveys

Information about program management was obtained from surveys administered for the original GAIN, PI, and NEWWS studies to staff members from each local office in the sample. Staff responses to these surveys provide rich information about local organizational conditions, how programs were implemented, interactions between their staff and clients, and interactions between their staff and supervisors. Responses to most of these questions were coded as five point or seven point Likert scales. In addition, staff background information was obtained from questions about their personal characteristics.

A comparison of staff survey dates with the client random assignment dates summarized in Table C1 indicates that the surveys were temporally aligned with the program participation of clients in the present analysis sample.[1] Thus, staff responses to these surveys reflect the conditions that prevailed when most of these clients were in the programs being studied.

For example, the GAIN staff surveys were administered in two waves between mid-1989 and mid-1991 (Riccio and Friedlander, 1992, pp. 17, 46, 190) and most experimental sample members in the present analysis were randomly assigned during 1989.[2] The PI staff surveys were administered in September and October 1991 (Kemple and Haimson, 1994, p. 32) and all PI experimental sample members in the present analysis were randomly assigned during the first three quarters of 1991. Lastly, the NEWWS staff surveys were administered between August and December 1993 (Freedman et al. 2000, p. 21) and most experimental sample members were randomly assigned in 1992 and 1993.

Both caseworkers ("frontline staff") and their unit supervisors were included in the staff surveys, and the local office sample size for each group is listed in Table C2. Completion rates for these surveys were uniformly high, exceeding 90 percent in most offices. Because frontline staff members are the main point of contact with clients, their responses were used to construct all but one of the program management measures.[3] For the remaining measure, responses from frontline staff *and* their supervisors were used to characterize their differences in perceptions.

---

[1] As noted in Section 3.2, restrictions that were necessary for the present analysis produced samples that differ in size from those used for the original studies of GAIN, PI, and NEWWS.

[2] Some GAIN staff were surveyed in both waves of the survey. Unique individuals accounted for 82 percent of all questionnaires completed by frontline staff, and 86 percent completed by unit supervisors.

C.1.2 The Measures

The present analysis developed six measures of program management that have been hypothesized to influence program impacts (Riccio, Bloom, and Hill, 2000). These include: three measures related to the service technology of each local program; one measure related to a specific program resource—frontline staff availability; one measure of the difference in views between frontline staff and their supervisors about their service technology; and one measure of the inconsistency in frontline staff views about their service technology.

C.1.2.1 Service Technology

Three constructs that were measured relate to key elements of a welfare-to-work program's service technology.[4] One construct reflects the employment message that its staff members convey to clients—that is, whether they encourage clients to take a job quickly, or to be more selective and take advantage of education and training opportunities first, in the hope of getting a better job later.

A second construct concerns the emphasis placed by each local program on providing personalized attention to its clients—that is, gaining an in-depth understanding of clients' personal histories and circumstances and trying to accommodate their individual needs and preferences when making assignments and referrals to specific program services and activities.

A third construct concerns how closely local staff monitor client participation in assigned program activities in order to keep abreast of their progress, their changing needs, and their involvement in the program.

To make each construct operational, a scale was developed for it using responses to the survey questions listed in Table C3. There were five goals in creating these scales. The first goal was to produce an *interpretable* measure, whose coefficient in the program impact model would be as meaningful as possible, given that the information represented by the scales has no natural metric. To help accomplish this, response values for each question were standardized to a mean of zero and a standard deviation of one, based on the individual responses of all frontline staff members in the sample. Then, a scale for each staff person was constructed by averaging her responses for the items in the scale. If a staff member did not respond to all items in a scale, her scale value was set to the mean of the items to which she did respond. Fortunately, as Table C4 illustrates, item response rates were typically quite high and thus, scale values for most staff respondents represent the mean for all items in the scale.

---

[3] In some cases, supervisors also see clients or fill in for frontline staff during vacations or illnesses. However, it was not possible to determine the extent of supervisors' interaction with clients.

[4] It is only possible to consider elements of local service technology that were addressed by the staff surveys.

A second goal for each scale was to *characterize program offices* in a way that is comparable across offices and controls for the possibility that different types of staff members perceive the same situation differently. To do so, a regression-adjusted office-level measure α_j, was created for each scale by estimating the following model:

$$Y_{ji} = \sum_{J} \alpha_j LO(J)_{ji} + \sum_{k} \beta_k X_{kji} + \varepsilon_{ji} \qquad \textbf{(C1)}$$

where

Y_{ji} = the scale value for frontline staff member i from local office j,
$LO(J)_{ji}$ = one if staff member i is from local office J and zero otherwise (with a separate indicator variable for each office J),
X_{kji} = personal characteristic k for frontline staff member i from office j (these variables are grand-mean centered), where the personal characteristics are:

- age of the staff member,
- whether she is female,
- whether she has formal education beyond a college degree,
- whether she had previous experience in a welfare-to-work program,
- whether she had received welfare in the past,

α_j = the mean scale value for office j, adjusted for its staff characteristics,
β_k = a regression coefficient indicating how the scale value varies with staff characteristic k,
ε_{ji} = a random error term for staff member i from local office j.

A third goal for each scale was to provide a *reliable* measure of the construct it is supposed to represent. The two indexes of reliability reported in Table C5 indicate that this goal was substantially achieved. The first index, "Cronbach's alpha," represents the *inter-item consistency* of each scale. Its possible values range from zero (total lack of consistency) to one (perfect consistency).[1] Estimated index values for the three service technology scales were 0.76, 0.83 and 0.84, indicating that they all have fairly high inter-item consistency.

The second index of reliability focuses on *inter-respondent consistency.* Its possible values also range from zero (total lack of consistency) to one (perfect consistency). Index values were estimated by a variance components analysis of the extent to which variation in mean office scale scores are due to within-office variation in staff scale values (lack of respondent consistency) versus between-office variation in the underlying construct.[6] The estimated values of this index are 0.76, 0.80 and 0.83 for the three service technology scales, indicating that they have fairly high inter-respondent consistency.

---

[1] Nunnally (1967), pp. 206-235 provides a detailed discussion of the reliability of a measure.

C.1.2.2 Staff Caseload Size

The average caseload size per frontline staff member is included in the present analysis to represent the hypothesis that large caseloads reduce the amount of time that caseworkers can spend with their clients or spend following up on their clients (Gueron and Pauly, 1991) and thereby reduce program performance.

The caseload size for each staff survey respondent was obtained from her answer to the following question: "How many clients are on your caseload today?" and the mean frontline staff response was used to represent the typical caseload size for each office. The average caseload size for the present sample of offices is 136 clients per frontline staff member; its standard deviation across offices is 67; and its range is 70 to 367.

C.1.2.3 Inconsistencies in Views Among Frontline Staff and Supervisors

The service technology scales described in section C.1.2.1 provide aggregate measures of each program's emphasis on quickly moving clients into jobs, on providing personalized client attention, and on closely monitoring client behavior and activities. In addition, they can be used to represent two other management-related factors that may have important effects on program performance.

First, the scales can provide information about the *inconsistency in views between frontline staff and unit supervisor* perceptions about these key elements of local service technology. Disparate views about these elements can reflect a disconnect in the local organizational hierarchy at a crucial point of service delivery. This disconnect may be due to unclear messages from supervisors or fundamental disagreements between them and their staff members. Regardless of its source, however, such a disconnect can undermine the quality of services provided to clients. Therefore, it is important to account for this factor when analyzing program performance. To do so, a scale of staff/supervisor differences in views was created as follows:

- The mean value for each of the three service technology scales was calculated separately for frontline staff and unit supervisors from each office,

- The absolute value was calculated of the *difference* between the mean staff and mean supervisor values for each scale for each office,

---

[6] The reliability coefficient "measures the ratio of the *true score* or parameter variance, relative to the *observed score* or total variance of [each office's sample mean outcome]. The reliability...will be close to 1 when (a) the group means...vary substantially across level-2 units (holding constant the sample size per group); or (b) the sample size ... is large" (Bryk and Raudenbush 1992, p. 40).

- A single office-level measure was obtained by summing the mean staff/supervisor differences for the three scales,[7]

- This sum was standardized to have a mean of zero and a standard deviation of one.

A second important construct that can be captured by the service technology scales is the *inconsistency in views among individual caseworkers* from a local program office. Organizational performance can suffer when staff members are divided—whether due to confusion or disagreement—over what the organization is or should be doing. Thus, it is often argued that managers' most important job is to instill a commonality of purpose, or "strong culture." To represent this construct (in the negative) a scale of caseworker variation in views was created as follows:

- The within-office variance of frontline staff responses was calculated for each of the three service technology scales,

- These three variances were summed for each office and the square root of the sum was computed,[8]

- The resulting measure was standardized to have a mean of zero and a standard deviation of one.

Table C6 summarizes the preceding management scales by listing the values for each by office plus their overall mean, standard deviation, and range across offices.

C.2 Constructing The Program Service Measures

Next consider how the program service measures were constructed.

C.2.1 The Data Source: Client Follow-up Surveys

As part of the original impact analyses for GAIN, PI, and NEWWS, a follow-up survey was administered to a random subsample of program and control group members from every local program office within two to three years after random assignment. The average size of the follow-up survey sample for an office was 258 persons; its standard deviation was 423 persons; and its range was 27 persons to 2,159 persons. Response rates for the survey ranged from 70 to 93 percent across the study counties.

Among the many issues addressed by this follow-up survey were a series of questions about employment and training services received by respondents during their first two years after random assignment. These questions were used for the present analysis, as described below, to construct office-level measures of the program/control group *differentials* in the receipt of three major types of services.[1]

---

[7] Five program offices had no information from unit supervisors. Thus, the overall sample mean scale values were imputed for these offices.

[8] It was not possible to calculate the within-office variance for the one program office with survey responses from only one line staff person. Thus, the overall sample mean scale value was imputed for this office.

C.2.2 The Program Service Measures

Two primary sets of issues arose in the specification of the program service measures: (1) what services to include and how to categorize them; and (2) how to characterize and compare services received by program group members and control group members.

With respect to the first issue, the primary goal of the present analysis was to capture the main streams of activities to which participants in welfare-to-work programs are exposed: basic education, job search assistance, and vocational training. Each of these general categories encompasses a range of specific activities. For example, basic education includes adult basic education, GED preparation, and English as a second language (ESL) classes; job search assistance includes both self-directed individual job search and participation in group job clubs; vocational training includes classroom training, on-the-job training, unpaid work experience, and post-secondary or vocational training. Clients may participate in none, one, or more than one of these types of activities.

With respect to the second issue, the primary goal of the present study was to represent accurately the increment in service receipt that was *caused by* the program. This was necessary in order to relate program-induced service receipt to program-induced earnings gains.

The best way to accomplish this task is to compare the services received by program group members to those received by control group members from each local program office. This service receipt *differential* provides a valid estimate of the difference between services that program members actually received, on average, and what they would have received on their own in the absence of the program—in other words, program-induced service receipt.

An office-level measure β_j of the service differential for each of the three types of services was obtained by estimating the following linear probability model from client follow-up survey data:

$$Y_{ji} = \sum_{J} \alpha_j LO(J)_{ji} + \sum_{J} \beta_j LO(J)_{ji} P_{ji} + \sum_{k} \delta_k CC_{kji} + \varepsilon_{ji} \qquad \textbf{(C2)}$$

where:

Y_{ji} = 100 if client i from office j received the service and zero otherwise,

---

[1] The service receipt rates and differentials in the present analysis may differ slightly from those in the original MDRC reports, due to differences in the samples used plus adjustments made by the original studies based on case file searches for some of the sites.

$LO(J)_{ji}$ = one if client i is from local office J and zero otherwise (with a separate indicator variable for each office J),

P_{ji} = one if client i from office j is a program group member and zero otherwise,

CC_{kji} = client characteristic k for client i from office j (each variable is grand-mean centered), where client characteristics are the same as those in Equation 1 of the program impact model,

α_j = the percentage of control group members from office j who received the service, controlling for client characteristics,

β_j = the program/control group service differential for office j,

δ_k = a regression coefficient for client characteristic k,

ε_{ji} = a random error term for client i from office j.

Table C7 lists the client survey sample size for each office and the values of its three service differential measures obtained by estimating the model represented in Equation C2.

C.3 Constructing the Measure of the Local Economic Environment

Lastly, consider how the measure of the local economic environment was constructed.

C.3.1 Calculation of Office Unemployment Rates

To characterize the local economic environment faced by experimental sample members during their two-year follow-up period, a measure of the local unemployment rate was created *for each local program office*. Information for measuring the unemployment rate was obtained from monthly, county-level data reported by the U.S. Bureau of Labor Statistics (BLS), Local Area Unemployment Statistics and the California Employment Development Department. Counties were used as the basis for measuring the unemployment rate faced by clients of each program office because they are a standard geographic unit and in many cases, they are the smallest unit for which unemployment rates are reported. Hence, they were judged to provide the best match for each office.

The measure was constructed in two steps. In the first step, the average unemployment rate faced by each sample member over her two-year follow-up period was calculated. For example, if an individual was randomly assigned in May 1991, then her average county unemployment rate was calculated from monthly data for July 1991 through June 1993.

An office-level average of individual-level average unemployment rates was then calculated as step two. Table C8 lists the average, the standard deviation, the minimum, and the maximum unemployment rates for clients from each office. The table also identifies the county within which each office is located.

C.3.2 An Alternative Unemployment Measure That Was Considered

An alternative reporting unit that was also considered, but not used, is the local "labor market area," or LMA. The Bureau of Labor Statistics, which reports monthly unemployment rates for LMAs, defines them as: "'an economically integrated geographic area within which individuals can reside and find employment within a reasonable distance or can readily change employment without changing their place of residence.' In addition, LMAs are nonoverlapping and geographically exhaustive" (BLS, http://www.bls.gov/laugeo.htm).

LMAs can include a single county, multiple counties, or a combination of other geographic units. To the extent that counties and LMAs differ, their unemployment rates may differ as well. Thus, to help assess the implications of choosing one geographic unit over the other, the correlation between their average unemployment rates for program offices was estimated. This correlation was 0.96, which suggests that the distinction between the two units was of little consequence for the present analysis.

C.3.3 An Additional Economic Indicator that Was Considered

At an early point in the study, an additional measure of the local economic environment was considered, but not used, because of concerns about its likely precision. This measure was based on estimates of the county-level job growth during the two-year follow-up period for sample members from each office. As was the case for unemployment rates, the job-growth measure was created in two steps; by first computing a value for each sample member and then averaging the values for all sample members from each office.

In the first step, the job growth rate for each sample member was computed from the employment levels reported for her county by the Bureau of Labor Statistics, Local Area Unemployment Statistics during the first and eighth quarters after random assignment. Because reported employment fluctuates from month to month due in part to sampling error, more stable measures for the beginning and end of the follow-up period were obtained by calculating the average employment level over five months, centered on the middle month of the beginning and ending calendar quarters. The employment growth rate for each sample member was then calculated as the annualized percentage change in the average employment level from the beginning to the end of her follow-up period. In step two, the average percentage growth rate for all sample members from each program office was computed to produce an office-level job growth measure.

Because the amount of job growth that occurs in a two-year period is likely to be small relative to the amount of sampling error in the estimates of local employment levels that "bracket" this period (especially for small areas, like counties) it was judged that the "signal-to-noise" ratio or reliability of the final measure was probably too low for it to be used in the analysis.[10]

---

[10] This measure is likely to be more stable over longer periods (e.g., a five-year follow-up period).

C.4 Assessing the Construct Validity of the Office-Level Measures Of Program Characteristics

Another issue to consider when assessing the "quality" of the preceding ten office-level measures is their *construct validity;* that is, the extent to which they represent the constructs they are intended to measure. A simple and effective way to make this assessment is to observe whether the estimated pattern of positive, negative and negligible correlations among the measures approximates the expected pattern of correlations for their hypothesized underlying constructs.

Table C9 facilitates this assessment by presenting the estimated correlation coefficients for all pairs of office-level measures, with the p-value for each coefficient in parentheses below it. When interpreting these findings, it should be noted that most social science measures contain substantial random error, and many social science constructs are only indirectly (and thus weakly) related to each other. Hence, one should not expect to see many strong correlations between the measures being assessed. Thus, one should rely on the *pattern* of correlations—not their absolute magnitudes—to judge whether a given measure is performing as it should if it is measuring the intended construct.

The first measure in the table—program emphasis on quick job entry for clients—is a four-item scale reflecting frontline staff perceptions about their programs. One would expect this characteristic of a program to be positively related to its reliance on job search assistance (which facilitates quick job entry) and negatively related to its reliance on vocational training (which delays client job entry) and on basic education (which delays client job entry and is not directly related to employment). Consistent with this expectation, the perceived emphasis on quick client job entry is: (1) positively correlated the program-induced increase in client receipt of job search assistance, and (2) negatively correlated with the corresponding increase in client receipt of basic education or vocational training.

The next three measures in the table—program emphasis on personalized client attention, closeness of client monitoring, and average caseload size—are arguably related to each other in predictable ways. One would expect that as caseload size increases, it becomes less possible for caseworkers to provide clients with personalized service or to monitor their activities closely. Thus, caseload size should be negatively correlated with the other two program characteristics, which, in turn, should be positively correlated with each other. As can be seen from the table, this is the case.

It is more difficult, however, to assess the last two management measures—inconsistency among staff and between staff and supervisors about service technology—because *a priori* expectations for their correlations with each other and with the other office-level measures are less clear. Thus, the findings in Table C9 do not present evidence for or against the construct validity of these two measures.

In general, this is also the case for the three service differential measures. The one exception, however, is the finding cited above that program focus on quick job entry is positively correlated with job search assistance and negatively correlated with basic education and vocational training. These correlations serve both to help validate the quick job entry measure (as noted above) and the measures of program reliance on the three types of employment and training services.

Lastly, consider the measure of unemployment rates, for which there also are not strong *a priori* expectations in terms of correlations with other measures in the table. Fortunately, however, there is an alternative basis for helping to validate this measure. Specifically, one would expect that control group earnings would be lower in areas with higher unemployment rates. And in fact, estimates of Equation 3 from the basic impact model indicate that this expected relationship exists.

In summary then, for the eight office-level measures whose construct validity could be assessed (at least in part) the present findings suggest that they are measuring what they were designed to measure.[11] For the other two measures, no assessment was possible.

C.5 Assessing the Statistical Significance of the Variation in Program Characteristics Across Offices

One final issue to consider when assessing the measures of program characteristics that are used as explanatory variables in the present model is whether they vary statistically significantly across offices. Table C10 indicates that this is the case for eight of the ten measures. For each measure, the significance level or p-value used to test the statistical significance of its across-office variation is listed.

The p-values for eight of the ten measures are 0.000, which means that they are statistically significant at beyond the 0.001-level. For a ninth measure—staff/supervisor inconsistency—the variation across offices was not statistically significant. This is probably because of the extremely small samples of *supervisors* per office. Nevertheless, the measure was included in the present model because it is the best available way to represent the construct of interest. For a tenth measure—frontline staff inconsistency—it was not clear how to test the statistical significance of its variation across offices. This is because it was constructed as a pooled standard deviation from three scales. Thus, there currently is no information about the statistical significance of the variation in this measure.

---

[11] Findings from the field research conducted for the original GAIN and NEWWS evaluations also support the face validity of the quick job entry, personalized attention, and closeness of monitoring measures based on the staff survey data. (Cross-validation of field research from PI is not available for the scales used in the current analysis.) The field research was able to document corresponding differences in actual staff practices across locations (and, in NEWWS, across treatment groups) that ranked differently on the relevant staff survey scales. For example, the observation that certain California counties in the GAIN evaluation had a higher or lower ranking than others on the quick job entry scale "made sense" when one considered the differences in their implementation strategies described by the field research for that evaluation.

Table C1

Random Assignment Dates and Sample Sizes for the Analysis Sample

Year and Quarter of Random Assignment	Number of Experimental Sample Members in the Analysis Sample	Percent of Experimental Sample Members in the Analysis Sample	Cumulative Percent of Experimental Sample Members in the Analysis Sample
GAIN			
1988			
Third Quarter	1,558	8.6	8.6
Fourth Quarter	2,458	13.6	22.2
1989			
First Quarter	2,859	15.8	37.9
Second Quarter	2,518	13.9	51.8
Third Quarter	4,396	24.3	76.1
Fourth Quarter	1,941	10.7	86.8
1990			
First Quarter	1,861	10.3	97.0
Second Quarter	535	3.0	100.0
PI			
1991			
First Quarter	2,090	48.6	48.6
Second Quarter	1,780	41.4	90.1
Third Quarter	426	9.9	100.0
NEWWS			
1991			
Second Quarter	100	0.2	0.2
Third Quarter	1,241	2.6	2.9
Fourth Quarter	2,986	6.4	9.2
1992			
First Quarter	4,235	9.0	18.2
Second Quarter	4,231	9.0	27.2
Third Quarter	4,589	9.8	37.0
Fourth Quarter	5,108	10.9	47.9
1993			
First Quarter	6,417	13.7	61.5
Second Quarter	5,808	12.4	73.9
Third Quarter	3,760	8.0	81.9
Fourth Quarter	2,977	6.3	88.2
1994			
First Quarter	2,538	5.4	93.6
Second Quarter	2,072	4.4	98.1
Third Quarter	891	1.9	99.9
Fourth Quarter	24	0.1	100.0

Table C2

Staff Survey Sample Sizes

Office	Number of Frontline Staff	Number of Supervisors
GAIN1[a]	12	3
GAIN2	31	4
GAIN3	46	9
GAIN4	49	7
GAIN5	27	4
GAIN6	26	4
GAIN7	42	6
GAIN8	50	7
GAIN9	15	2
GAIN10	40	4
GAIN11	54	8
GAIN12	30	4
GAIN13	48	7
GAIN14	20	2
GAIN15	59	9
GAIN16	10	2
GAIN17	83	14
GAIN18	50	9
GAIN19	9	2
GAIN20	41	6
GAIN21	24	3
GAIN22	10	2
PI1	2	1
PI2	8	1
PI3	7	1
PI4	3	1
PI5	4	1
PI6	3	1
PI7	7	1
PI8	8	1
PI9	7	1
PI10	8	1

(continued)

Table C2 (continued)

Staff Survey Sample Sizes

Office	Number of Frontline Staff	Number of Supervisors
NEWWS1	4	0
NEWWS2	2	0
NEWWS3	14	1
NEWWS4	6	1
NEWWS5	16	4
NEWWS6	17	0
NEWWS7	9	2
NEWWS8	1	1
NEWWS9	11	2
NEWWS10	7	1
NEWWS11	3	0
NEWWS12	10	1
NEWWS13	3	2
NEWWS14	21	2
NEWWS15	4	1
NEWWS16	8	1
NEWWS17	40	8
NEWWS18	19	3
NEWWS19	14	2
NEWWS20	3	1
NEWWS21	7	2
NEWWS22	61	10
NEWWS23	18	4
NEWWS24	30	6
NEWWS25	38	7
NEWWS26	24	4
NEWWS27	2	0
Average	21	3
Standard Deviation	19	3
Range	1 to 83	0 to 14

Note:

a. The GAIN staff surveys were administered in two waves (Riccio and Friedlander 1992, pp. 17, 46, 190). The sample within each GAIN office is roughly evenly split between the first and second waves. Unique individuals accounted for 82 percent of all questionnaires completed by frontline staff, and 86 percent completed by unit supervisors.

Table C3: Scales and Survey Items for the Service Technology Measures

Scale and Items	Response Scale
Emphasis on Moving Clients Into Jobs Quickly	
• Based on the practices in your unit, what would you say is the more important goal of your unit: to help clients get jobs as quickly as possible or to raise the education or skill levels of clients so that they can get jobs in the future?	1 .... 7 skills. jobs
• In your opinion, which should be the more important goal of your unit: to help clients get jobs as quickly as possible or to raise the education or skill levels of clients so that they can get jobs in the future?	1 .... 7 skills jobs
• After a short time in the program, an average welfare mother is offered a low-skill, low-paying job that would make her slightly better off financially. Assume she has two choices: either to take the job and leave welfare OR to stay on welfare and wait for a better opportunity. If you were asked, what would your personal advice to this client be?	1 .... 7 welfare jobs
• What advice would your supervisor want you to give to a client of this type?	1 .... 7 welfare jobs
Emphasis on Personalized Client Attention	
• In our program, there is more emphasis on the number of clients served than on the quality of services.	1 .... 7 strongly agree strongly disagree
• Do you feel that in your unit not enough time or enough time is being spent with clients during the intake process?	1 .... 7 not enough enough
• During intake, how much effort does the staff make to learn about the client's family problems in depth?	1 .... 7 very little a great deal
• During intake, how much effort does the staff make to learn about the client's goals and motivation to work in depth?	1 .... 7 very little a great deal
• In your opinion, how well is the program tailoring the educational, training and work experience services that clients receive to their particular needs, circumstances, and goals?	1 .... 7 very poorly very well
Closeness of Client Monitoring	
• How closely would you say the staff of your unit is monitoring clients?	1 .... 7 not very very
• Suppose a client has been assigned to Adult Basic Education (ABE, GED, ESL) but has not attended it at all. How long would is usually take for staff to learn about this situation from the service provider?	1 .... 5 1 or fewer weeks 5 or more weeks
• Suppose a client has been assigned to vocational education but has not attended it at all. How long would is usually take for staff to learn about this situation from the service provider?	1 .... 5 1 or fewer weeks 5 or more weeks
• Suppose a client has a part-time job that deferred her from other program obligations. How closely would you say your agency is monitoring whether clients quit or lose part-time jobs?	1 .... 7 not very very
• Once your agency learned that a client lost or quit a part-time job, how long on average would it take before the client was assigned to another program component?	1 .... 8 1 or fewer weeks 8 or more weeks

BEST COPY AVAILABLE

Table C4

Item Response Patterns for the Service Technology Scales

	Percent of all frontline staff who provided responses to:					
Scale[a]	**5 items**	**4 items**	**3 items**	**2 items**	**1 item**	**0 items**
Emphasis on moving clients into jobs quickly	n.a.	85.6	8.6	3.6	0.2	2.0
Emphasis on personalized client attention	84.7	7.0	1.5	2.4	2.0	2.4
Closeness of client monitoring	59.8	21.0	7.0	6.1	2.0	4.0

Note:

a. See Table C3 for a description of each item in these scales.

Table C5

Reliability Assessments for the Service Technology Scales

	Service Technology Scale		
	Emphasis on Moving Clients into Jobs Quickly	**Emphasis on Personalized Client Attention**	**Closeness of Client Monitoring**
Internal consistency: Cronbach's alpha[a]	0.84	0.83	0.76
Inter-rater reliability[b]	0.83	0.76	0.80

Notes:

a. Cronbach's alpha was estimated from the office-level means of each of the items in the scale.

b. Inter-rater reliability was estimated from the scale value for each staff survey respondent using a two-level hierarchical model.

Table C6

Values of the Program Management Measures For Each Local Program Office

	Emphasis on Moving Clients into Jobs Quickly	Emphasis on Personalized Client Attention	Closeness of Client Monitoring	Staff Caseload Size	Frontline Staff Supervisor Inconsistency About Service Technology	Frontline Staff Inconsistency About Service Technology
GAIN1	1.59	0.01	0.23	70	0.50	0.01
GAIN2	0.89	-0.79	0.10	74	-0.10	-0.04
GAIN3	0.35	0.47	-0.18	70	-1.08	-0.03
GAIN4	1.22	-0.31	0.35	93	-0.08	0.14
GAIN5	0.72	0.04	0.56	71	-0.16	-0.10
GAIN6	-0.76	1.27	0.44	106	-1.01	-0.16
GAIN7	-0.05	0.40	0.28	102	-1.17	0.19
GAIN8	-0.68	0.80	0.35	109	-0.58	0.14
GAIN9	-0.63	1.33	0.91	109	0.43	-0.65
GAIN10	-1.05	0.71	0.19	116	-0.15	0.53
GAIN11	-1.36	1.12	-1.33	75	-1.34	0.05
GAIN12	-0.45	-1.03	0.46	131	0.23	-0.54
GAIN13	-0.68	-0.36	0.07	143	0.30	0.37
GAIN14	-1.28	1.88	0.70	85	-0.52	-0.34
GAIN15	-0.61	0.13	0.04	137	-0.86	0.14
GAIN16	0.03	1.86	0.73	99	-0.40	-1.34
GAIN17	-0.73	-0.22	0.25	103	-1.25	-0.01
GAIN18	-0.84	0.33	0.42	103	-0.59	-0.04
GAIN19	-0.91	1.80	0.73	132	0.70	-0.99
GAIN20	-1.34	-0.26	-0.35	135	-0.37	-0.28
GAIN21	-1.16	1.21	0.70	90	-0.32	-0.30
GAIN22	-0.78	0.47	-0.16	124	1.98	-0.22
PI1	-0.32	0.03	0.41	98	1.30	4.54
PI2	-0.50	2.04	1.89	250	-0.53	0.83
PI3	0.33	-0.20	0.34	135	-0.31	0.60
PI4	1.14	-1.53	-1.65	222	1.23	-0.66
PI5	0.08	-0.66	-0.60	326	0.11	3.30
PI6	0.01	-1.22	-2.12	367	0.52	0.13
PI7	0.22	2.29	-0.33	298	3.14	-0.73
PI8	-1.67	-0.30	0.18	130	1.38	0.62
PI9	0.01	-0.04	0.37	98	1.45	0.53
PI10	0.61	-1.21	0.60	153	1.98	-0.27

(continued)

Table C6 (continued)

Values of the Program Management Measures For Each Local Program Office

	Emphasis on Moving Clients into Jobs Quickly	Emphasis on Personalized Client Attention	Closeness of Client Monitoring	Staff Caseload Size	Frontline Staff Supervisor Inconsistency About Service Technology	Frontline Staff Inconsistency About Service Technology
NEWWS1	0.59	-0.40	-2.52	113	0.00	0.29
NEWWS2	-0.74	0.29	0.90	95	0.00	-0.24
NEWWS3	0.60	0.04	0.85	74	-0.95	0.04
NEWWS4	2.07	0.45	1.11	108	-0.63	-2.13
NEWWS5	1.73	-1.21	0.48	102	-0.81	-0.45
NEWWS6	-0.15	0.05	0.30	105	0.00	0.50
NEWWS7	0.38	-0.52	0.61	90	-0.39	0.95
NEWWS8	1.29	2.11	1.83	100	3.20	0.00
NEWWS9	2.50	-0.01	0.75	104	0.01	-0.60
NEWWS10	0.87	-0.40	0.25	122	0.47	1.55
NEWWS11	0.75	-0.96	-0.87	100	0.00	-1.36
NEWWS12	0.30	-0.04	-0.30	95	0.29	-0.56
NEWWS13	1.36	0.02	1.27	77	0.20	0.00
NEWWS14	0.55	-1.22	0.76	120	-0.17	-0.35
NEWWS15	1.97	-1.02	0.43	163	1.28	-1.11
NEWWS16	-0.62	-2.00	-2.81	118	0.22	-1.69
NEWWS17	-0.49	-1.11	-1.41	258	-0.68	0.78
NEWWS18	-0.37	-0.29	-0.57	167	-1.49	-0.40
NEWWS19	-1.02	0.60	-0.46	88	-1.27	-0.67
NEWWS20	0.74	1.13	0.81	128	-1.13	0.12
NEWWS21	-0.89	-1.26	-1.66	260	-0.32	-0.07
NEWWS22	-0.87	-0.68	-1.18	138	-0.96	0.03
NEWWS23	-1.04	-0.68	-2.04	293	0.46	0.79
NEWWS24	-0.88	-0.46	-0.55	200	-0.77	0.30
NEWWS25	-0.91	-0.90	-0.98	169	-0.60	0.42
NEWWS26	-0.94	-0.28	-0.75	162	-0.41	-0.11
NEWWS27	1.77	-1.32	1.22	115	0.00	-1.44
Average	0	0	0	136	0	0
Standard Deviation	1	1	1	67	1	1
Range	-1.7 to 2.5	-2.0 to 2.3	-2.8 to 1.9	70 to 367	-1.5 to 3.2	-2.1 to 4.5

Table C7

Sample Sizes and Values of the Service Differential Measures For Each Local Program Office

		Service Differential for:		
Office	**Number of respondents to client survey**	**Basic Education**	**Job Search**	**Vocational Training**
GAIN1	56	3.5	15.5	14.3
GAIN2	313	8.3	11.3	-11.0
GAIN3	36	-4.2	-13.1	-10.3
GAIN4	396	3	17.9	-10.4
GAIN5	131	5.3	7.7	15.0
GAIN6	55	18.2	35.7	14.7
GAIN7	104	19.3	6.6	-2.2
GAIN8	94	8.1	20.2	3.4
GAIN9	37	9.6	-1.2	16.0
GAIN10	94	11	16.1	3.1
GAIN11	656	26.4	11.2	5.5
GAIN12	62	21	8.3	4.1
GAIN13	39	-11	5.9	11.9
GAIN14	152	14.5	8.7	12.3
GAIN15	157	33	6.1	2.0
GAIN16	98	23.6	11.5	17.7
GAIN17	106	11.7	2.7	9.9
GAIN18	146	24.7	5.1	7.9
GAIN19	86	4.2	25.8	14.0
GAIN20	67	22	0.1	-3.2
GAIN21	122	23.2	21.9	24.7
GAIN22	156	20.4	15	3.3
PI1	70	0.1	30.2	9.0
PI2	65	18.8	31	14.7
PI3	55	-2.9	15.6	2.8
PI4	54	-7.3	22	-4.8
PI5	74	-7.5	13.5	-5.9
PI6	77	-1	6.3	-0.3
PI7	68	28.4	23.7	8.4
PI8	64	15.3	21.3	-5.3
PI9	74	-1.7	27.7	6.8
PI10	91	7.8	13.2	4.2

Table C7 (continued)

Sample Sizes and Values of the Service Differential Measures For Each Local Program Office

		Service Differential for:		
Office	**Number of respondents to client survey**	**Basic Education**	**Job Search**	**Vocational Training**
NEWWS1	31	-7.7	18.4	-16.5
NEWWS2	31	20.9	44.1	34.6
NEWWS3	110	6.9	34.6	22.6
NEWWS4	313	2.5	28.8	-1.7
NEWWS5	742	-0.3	31.8	0.1
NEWWS6	298	4.9	31.3	7.6
NEWWS7	53	-8.8	47.3	21.6
NEWWS8	30	7.4	34.6	-8.6
NEWWS9	268	-1.7	32.7	-1.1
NEWWS10	255	3.1	35.2	11.6
NEWWS11	27	6.3	11.1	18.9
NEWWS12	1,853	6.6	30.4	7.0
NEWWS13	264	28.4	26.7	5.1
NEWWS14	1,668	6.5	20.9	3.2
NEWWS15	614	39.3	19.1	3.8
NEWWS16	244	-1.4	7.9	2.6
NEWWS17	675	11.7	9.9	11.5
NEWWS18	681	12	11.4	9.6
NEWWS19	2,159	19.1	10.3	10.6
NEWWS20	205	49.6	10.7	-5.3
NEWWS21	173	3.3	5.5	14.4
NEWWS22	165	7.6	0.7	4.6
NEWWS23	61	11.9	5.9	-21.0
NEWWS24	84	16.6	17.6	9.4
NEWWS25	124	11.2	-1.9	7.2
NEWWS26	55	0.1	17.6	3.8
NEWWS27	197	40.3	18.1	-2.3
Average	258	10.9	17.0	5.5
Standard Deviation	423	12.8	12.2	10.2
Range	27 to 2,159	-11.0 to 49.6	-13.1 to 47.3	-21.0 to 34.6

Table C8
Descriptive Statistics for the Unemployment Rate At Each Local Program Office

Office	County	Average	Standard Deviation	Minimum	Maximum
GAIN1	COUNTY4	8.10	0.53	7.24	9.00
GAIN2	COUNTY4	7.42	0.68	6.58	9.00
GAIN3	COUNTY3	4.58	0.36	4.10	5.15
GAIN4	COUNTY4	7.60	0.72	6.58	9.00
GAIN5	COUNTY4	7.42	0.78	6.58	9.00
GAIN6	COUNTY3	4.56	0.37	4.10	5.15
GAIN7	COUNTY4	4.59	0.35	4.10	5.15
GAIN8	COUNTY3	4.59	0.36	4.10	5.15
GAIN9	COUNTY3	4.60	0.35	4.10	5.15
GAIN10	COUNTY3	4.63	0.36	4.10	5.15
GAIN11	COUNTY5	4.86	0.31	4.48	5.33
GAIN12	COUNTY2	6.89	0.34	6.65	7.50
GAIN13	COUNTY2	6.92	0.35	6.65	7.50
GAIN14	COUNTY1	14.05	1.26	12.35	15.79
GAIN15	COUNTY2	6.81	0.29	6.65	7.50
GAIN16	COUNTY1	13.93	1.23	12.35	15.79
GAIN17	COUNTY2	6.98	0.34	6.65	7.50
GAIN18	COUNTY3	4.59	0.34	4.10	5.15
GAIN19	COUNTY1	13.92	1.24	12.35	15.79
GAIN20	COUNTY2	7.01	0.36	6.65	7.50
GAIN21	COUNTY1	13.84	1.21	12.35	15.79
GAIN22	COUNTY1	14.29	1.16	12.35	15.79
PI1	COUNTY14	6.81	0.02	6.79	6.83
PI2	COUNTY9	9.84	0.12	9.58	9.94
PI3	COUNTY13	8.06	0.07	7.89	8.11
PI4	COUNTY12	8.68	0.14	8.59	9.08
PI5	COUNTY11	7.15	0.03	7.12	7.18
PI6	COUNTY10	7.03	0.07	6.90	7.10
PI7	COUNTY9	9.83	0.12	9.58	9.94
PI8	COUNTY8	6.40	0.02	6.38	6.42
PI9	COUNTY7	6.73	0.03	6.70	6.80
PI10	COUNTY6	6.47	0.05	6.35	6.51

(continued)

Table C8 (continued)

Descriptive Statistics for the Unemployment Rate At Each Local Program Office

Office	County	Average	Standard Deviation	Minimum	Maximum
NEWWS1	COUNTY23	4.46	0.06	4.38	4.55
NEWWS2	COUNTY23	4.45	0.07	4.38	4.55
NEWWS3	COUNTY24	3.56	0.28	3.31	4.18
NEWWS4	COUNTY15	11.35	0.43	10.60	11.80
NEWWS5	COUNTY15	11.40	0.42	10.60	11.80
NEWWS6	COUNTY23	4.96	0.25	4.51	5.35
NEWWS7	COUNTY23	4.46	0.08	4.38	4.68
NEWWS8	COUNTY23	4.46	0.07	4.38	4.68
NEWWS9	COUNTY15	11.37	0.42	10.60	11.80
NEWWS10	COUNTY15	11.35	0.41	10.60	11.80
NEWWS11	COUNTY23	4.45	0.07	4.38	4.68
NEWWS12	COUNTY20	6.07	0.35	5.48	6.67
NEWWS13	COUNTY15	11.35	0.43	10.60	11.80
NEWWS14	COUNTY22	5.19	0.84	3.91	6.78
NEWWS15	COUNTY15	11.39	0.43	10.60	11.80
NEWWS16	COUNTY19	6.77	0.73	5.93	8.52
NEWWS17	COUNTY21	3.62	0.40	2.95	4.32
NEWWS18	COUNTY21	3.63	0.40	2.95	4.32
NEWWS19	COUNTY20	6.07	0.35	5.48	6.67
NEWWS20	COUNTY15	11.37	0.42	10.60	11.80
NEWWS21	COUNTY19	6.91	0.79	5.93	8.52
NEWWS22	COUNTY17	5.33	0.15	4.99	5.51
NEWWS23	COUNTY18	3.51	0.04	3.40	3.62
NEWWS24	COUNTY17	5.33	0.15	4.99	5.51
NEWWS25	COUNTY17	5.33	0.15	4.99	5.51
NEWWS26	COUNTY16	5.73	0.12	5.47	5.95
NEWWS27	COUNTY15	11.33	0.40	10.60	11.80
Average		7.36			
Standard Deviation		3.10			
Range		3.51 to 14.29			

Table C9

Correlations Among Measures of Program Management, Program Services, and the Economic Environment

(Correlations are followed by p-values in parentheses)

	Emphasis on Moving Clients into Jobs Quickly	Emphasis on Personalized Client Attention	Closeness of Client Monitoring	Staff Caseload Size	Frontline Staff/ Supervisor Inconsistency About Service Technology	Frontline Staff Inconsistency About Service Technology	Differential in Basic Education Participation	Differential in Job Search Participation	Differential in Vocational Training Participation	Unemployment Rate
Emphasis on Moving Clients into Jobs Quickly	1.000	-0.209	0.296	-0.158	0.210	-0.208	-0.111	0.341	-0.238	0.278
	(0.000)	(0.112)	(0.023)	(0.232)	(0.111)	(0.115)	(0.404)	(0.008)	(0.069)	(0.033)
Emphasis on Personalized Client Attention	-0.209	1.000	0.500	-0.206	0.064	-0.035	0.309	0.169	0.239	0.258
	(0.112)	(0.000)	(0.000)	(0.118)	(0.629)	(0.791)	(0.017)	(0.200)	(0.069)	(0.049)
Closeness of Client Monitoring	0.296	0.500	1.000	-0.455	0.083	-0.028	0.290	0.424	0.285	0.341
	(0.023)	(0.000)	(0.000)	(0.000)	(0.530)	(0.833)	(0.026)	(0.001)	(0.029)	(0.008)
Staff Caseload Size	-0.158	-0.206	-0.455	1.000	0.211	0.220	-0.062	-0.163	-0.187	-0.057
	(0.232)	(0.118)	(0.000)	(0.000)	(0.109)	(0.094)	(0.638)	(0.216)	(0.155)	(0.668)
Frontline Staff/ Supervisor Inconsistency About Service Technology	0.210	0.064	0.083	0.211	1.000	0.064	-0.089	0.268	-0.146	0.198
	(0.111)	(0.629)	(0.530)	(0.109)	(0.000)	(0.632)	(0.504)	(0.040)	(0.270)	(0.133)
Frontline Staff Inconsistency About Service Technology	-0.208	-0.035	-0.028	0.220	0.064	1.000	-0.266	0.120	-0.075	-0.245
	(0.115)	(0.791)	(0.833)	(0.094)	(0.632)	(0.000)	(0.042)	(0.363)	(0.570)	(0.061)
Differential in Basic Education Participation	-0.111	0.309	0.290	-0.062	-0.089	-0.266	1.000	-0.111	0.085	0.269
	(0.404)	(0.017)	(0.026)	(0.638)	(0.504)	(0.042)	(0.000)	(0.404)	(0.522)	(0.040)
Differential in Job Search Participation	0.341	0.169	0.424	-0.163	0.268	0.120	-0.111	1.000	0.288	0.148
	(0.008)	(0.200)	(0.001)	(0.216)	(0.040)	(0.363)	(0.404)	(0.000)	(0.027)	(0.265)
Differential in Vocational Training Participation	-0.238	0.239	0.285	-0.187	-0.146	-0.075	0.085	0.288	1.000	0.065
	(0.069)	(0.069)	(0.029)	(0.155)	(0.270)	(0.570)	(0.522)	(0.027)	(0.000)	(0.625)
Unemployment Rate	0.278	0.258	0.341	-0.057	0.198	-0.245	0.269	0.148	0.065	1.000
	(0.033)	(0.049)	(0.008)	(0.668)	(0.133)	(0.061)	(0.040)	(0.265)	(0.625)	(0.000)

Table C10

Statistical Significance of the Variation in Program Characteristics Across Offices

Across-Office Variation in:	Statistical Significance (p-value)
Emphasis on Moving Clients into Jobs Quickly	0.000
Emphasis on Personalized Client Attention	0.000
Closeness of Client Monitoring	0.000
Staff Caseload Size	0.000
Frontline Staff/Supervisor Inconsistency About Service Technology	>0.500
Frontline Staff Inconsistency About Service Technology	*a*
Differential in Basic Education Participation	0.000
Differential in Job Search Participation	0.000
Differential in Vocational Training Participation	0.000
Unemployment Rate	0.000

Note:

a: It was not possible to compute the statistical significance of the variation across offices for this measure.

Appendix D

Testing the Sensitivity of Findings from the Impact Model

This appendix tests the sensitivity of key findings from the program impact model. For this purpose, the model was re-estimated for samples that varied systematically according to four strategies: (1) deleting the Riverside GAIN offices and the Portland NEWWS offices separately and jointly; (2) deleting a progressively increasing number of offices with the most extreme positive and negative impact estimates; (3) deleting 17 program offices that administered random assignment at the point of welfare application or redetermination instead of when sample members attended a GAIN, PI, or NEWWS orientation; and (4) deleting a progressively increasing number of offices with the most extreme positive and negative values of selected independent variables.

The first strategy deletes the four Riverside[1] GAIN offices and the seven Portland NEWWS offices because these sites produced the largest program impacts in GAIN and NEWWS, respectively (Riccio, et al., 1994 and Freedman, et al., 2000). Thus, it is possible that their unusually high performance had a disproportionate influence on the present findings. The second strategy takes account of the fact that not all high-performing offices are from Riverside GAIN or Portland NEWWS. In addition, it accounts for the possibility that offices with the most extreme negative impacts may be driving the present results. The third strategy deletes 17 offices that administered random assignment early during intake because their experimental design may have created program impact estimates that were systematically smaller than those of the other offices. The fourth strategy deletes offices with the highest and lowest values of selected independent variables to account for the possibility that these offices drive the coefficient estimates.

This appendix first examines the sensitivity of coefficient estimates for program characteristics and then for client characteristics. The results indicate that both sets of coefficient estimates are quite robust.

D.1 Sensitivity of the Estimated Relationships Between *Program Characteristics* and Program Impacts

D.1.1 Deleting Riverside GAIN and Portland NEWWS

Table D1 presents sensitivity tests of the estimated relationships between program characteristics and program impacts produced by systematically deleting Riverside GAIN and Portland NEWWS offices from the sample. The first column in the table repeats the original full-sample results reported in Table 8. This serves as a point of departure. Column two lists the results obtained when the four Riverside GAIN offices were deleted from the sample (leaving 55 offices); column three lists the results obtained when the seven Portland NEWWS offices were deleted (leaving 52 offices); column four lists the results obtained when *both* Riverside GAIN and Portland NEWWS offices were deleted (leaving 48 offices).

---

[1] The NEWWS study also included offices in Riverside County, California, which are considered as different offices from those in the Riverside GAIN program.

The last rows in the table list the overall mean impact, the mean counterfactual and the mean impact as a percentage of the mean counterfactual for each sample of program offices. Note that the percentage impact drops from 18.0 percent for the full sample to 11.6 percent without the high-performing Riverside GAIN and Portland NEWWS offices.

Reading across each row makes it possible to check the sensitivity of coefficient estimates for each program characteristic. For example, the full-sample coefficient estimate for "emphasis on moving clients into jobs quickly" is $720, with a p-value of 0.000. When the Riverside GAIN offices are deleted, it becomes $556 with a p-value of 0.001. When instead the Portland NEWWS offices are deleted, it is $740, with a p-value of 0.000. Finally, when both Riverside GAIN *and* Portland NEWWS are removed, the estimate is $525, with a p-value of 0.004. Thus, the basic finding is quite robust.

Similar patterns hold for most of the other program characteristics. Although coefficient estimates change somewhat as offices are deleted, they tend to remain statistically significant and support the same basic conclusion. For three variables—caseload size, the basic education service differential, and the unemployment rate—the coefficient is statistically significant when *either* Riverside GAIN or Portland NEWWS is deleted, but drops below the 0.10-level when both sites are removed. This may be due to the loss of statistical power produced by deleting roughly one fifth of the offices in the full sample.

D.1.2 Deleting Positive and Negative Outlier Offices based on Impact Estimates

Table D2 presents sensitivity test results for program characteristics produced by systematically removing offices with the highest and lowest impact estimates. As in Table D1, the first column repeats the original full-sample results. The second column reports results for a subsample without the offices that had the two most positive and negative impact estimates. The third column deletes an additional office on each end of the impact distribution, and so on, until the last column, which deletes a total of ten offices.

The bottom row of the table indicates that the overall average impact as a percentage of the average counterfactual remains stable at 17 to 18 percent throughout the deletion process. This suggests an approximate balance in the positive and negative impact estimates that were deleted.

Once again, the sensitivity of the coefficient estimate for each program characteristic can be tested by reading across the row for that characteristic. And once again, the story is roughly the same: although coefficient estimates vary somewhat with the sample, their signs do not change and most coefficients maintain their statistical significance.

D.1.3 Deleting Offices with Early Random Assignment

A third series of sensitivity tests was conducted by deleting the 17 program offices (ten from PI and seven from NEWWS) that administered random assignment early in the sample intake process—at the point of welfare application or redetermination—instead of later in the process when sample members attended a GAIN, PI, or NEWWS orientation (which was the point of random assignment for the other 42 offices). Because these 17 offices administered random assignment early during intake, they had a greater margin for "fall-off" between random assignment to the program group and program participation. Hence, their experimental design may have diluted the program/control group treatment contrast and thereby created program impact estimates that were systematically smaller than those of the other offices. Yet it is also possible that participation mandates may generate impacts even prior to program orientation through what is commonly referred to as a "deterrent effect." Informing people that their eligibility for a full welfare grant would henceforth be contingent upon their enrollment in a welfare-to-work program may, in and of itself, encourage some to seek employment on their own instead of enrolling in the program. In evaluations where random assignment occurs at orientation, such deterrent effects are not captured by the reported impact estimates.

Table D3 presents the results of re-estimating the coefficients for program characteristics from data for this restricted sample. As can be seen from the bottom panel of the table, deleting the 17 offices *increased* the overall average impact substantially from $879 (18.0 percent of the counterfactual) for the full sample to $1,134 (23.6 percent) for the restricted sample. Thus, overall, early random assignment was related to smaller program impacts.

Nevertheless, the basic relationships between program characteristics and program impacts were still apparent in the results for the restricted sample. As the table indicates, although coefficient estimates vary somewhat, their signs do not change and most coefficients maintain their statistical significance. The fact that significance levels decline overall is probably due largely to the one-quarter reduction in the number of sample offices.

D.1.4 Deleting Offices Based on High and Low Values of Selected Independent Variables

A fourth series of sensitivity tests is presented in Table D4. This strategy systematically removed offices with the highest and lowest values of five office-level independent variables. These five variables were selected because their coefficients in the impact model were the largest in magnitude and were the most statistically significant of the program characteristics reported in Table 8.

Like the other tables in this appendix, the first column repeats the original full-sample results, and reading across each row makes it possible to check the sensitivity of coefficient estimates for the selected program characteristic. The second column reports results for a subsample without the offices that had the two most positive and negative values of *the particular program characteristic listed in each row of the table*. The third column deletes an additional office on each end of the program characteristic distribution, and so on, until the last column, which deletes a total of ten offices.

Thus, unlike the other tables in this appendix, reading *down* a column in Table D4 is not meaningful: the entry in each cell represents the coefficient estimate for the indicated program characteristic when offices with the highest and lowest values for the program characteristic listed in that row were deleted from the sample. For example, the first row shows that after deleting the offices with the two highest and two lowest values for the scale "emphasis on moving clients into jobs quickly," the estimated coefficient for this variable is $881 with a p-value of 0.000. The remaining entries in this row show that the coefficient estimate remains robust to deletion of additional offices on either end of the distribution *for this scale*.

Overall, Table D4 shows that the strong results for the five program characteristic variables remain robust to deletion of "outliers" based on each characteristic, with the exception of the service differential for basic education. For this variable, the magnitude of the effect remains similar to the full-sample estimate, but the statistical significance grows weaker with the deletion of additional offices.

D.2 Sensitivity of the Estimated Relationships between *Client Characteristics* and Program Impacts

D.2.1 Deleting Riverside GAIN and Portland NEWWS

Table D5 presents sensitivity tests of the relationships between client characteristics and program impacts produced by systematically deleting Riverside GAIN and Portland NEWWS, as described above. Coefficients for these client characteristics were estimated simultaneously with those for the program characteristics examined in Table D1.

Reading across each row of Table D5, one can see that the results for client characteristics are quite robust. In most cases, the coefficient estimates and their statistical significance change very little across subsamples, especially coefficient estimates that were highly significant for the full sample.

D.2.2 Deleting Positive and Negative Outlier Offices based on Impact Estimates

Table D6 presents sensitivity tests of the relationships between client characteristics and program impacts produced by deleting progressively more offices with the highest and lowest impacts estimates. These findings for client characteristics were estimated simultaneously with those for program characteristics reported in Table D2. The findings in Table D6 indicate no notable changes in either the magnitudes of the estimated coefficients, or in their statistical significance.

D.2.3 Deleting Offices with Early Random Assignment

Table D7 presents sensitivity tests of the relationships between client characteristics and program impacts produced by deleting offices that conducted early random assignment. Coefficients for these client characteristics were estimated simultaneously with those for the program characteristics examined in Table D3. For client characteristics that were statistically significant using the full sample, the coefficient estimates remain approximately the same magnitude and maintain their statistical significance in the restricted sample. Some other coefficients switch signs; all but one, however, still do not attain statistical significance.

D.3 Conclusions

Tables D1 through D7 present strong evidence that most of the key findings for the present study (especially those that are highly statistically significant for the full sample) are robust to the deletion of program offices with the most extreme impact estimates, with an especially early point of random assignment, or with the most extreme values of the independent variables. Thus, the findings—and the conclusions they suggest—appear not to be limited to a few idiosyncratic sites. Instead, they represent a fairly pervasive phenomenon that might generalize to a broad range of welfare-to-work programs in other settings.

Table D1

Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Riverside GAIN and Portland NEWWS

Program Characteristic	Full sample (n=59 offices)		Without Riverside GAIN (n=55 offices)		Without Portland NEWWS (n=52 offices)		Without Riverside GAIN *and* Without Portland NEWWS (n=48 offices)	
	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)[b]
Program Management								
Emphasis on moving clients into jobs quickly	720	0.000	556	0.001	740	0.000	525	0.004
Emphasis on personalized client attention	428	0.000	359	0.001	455	0.000	334	0.016
Closeness of client monitoring	-197	0.110	-221	0.048	-178	0.194	-173	0.231
Staff caseload size	-4	0.003	-2	0.041	-4	0.010	-2	0.311
Frontline staff/supervisor inconsistency about service technology	-159	0.102	-159	0.070	-131	0.251	-183	0.110
Frontline staff inconsistency about service technology	124	0.141	80	0.278	156	0.072	86	0.462
Service Differential								
Basic education	-16	0.017	-12	0.064	-19	0.006	-14	0.101
Job search assistance	1	0.899	15	0.167	-10	0.339	7	0.572
Vocational training	7	0.503	10	0.240	-5	0.700	-1	0.966
Economic Environment								
Unemployment rate	-94	0.004	-79	0.014	-68	0.061	-45	0.283
Mean Program Impact on Earnings	879	0.000	700	0.000	765	0.000	565	0.000
Mean Counterfactual	4,871	0.000	4,946	0.000	4,794	0.000	4,861	0.000
Impact as Percent of Counterfactual	18.0		14.2		16.0		11.6	

Notes:

[a] Coefficient values are in 1996 dollars.

[b] Unlike values reported in most other tables in this paper, the p-values listed here are not computed from robust standard errors due to the small sample size at Level Two that resulted from the restriction imposed in this sensitivity test.

Table D2

Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Offices with the Highest and Lowest Impact Estimates

Program Characteristic	Full Sample (n=59 offices)		Without offices with 2 highest and 2 lowest impacts (n=55 offices)		Without offices with 3 highest and 3 lowest impacts (n=53 offices)		Without offices with 4 highest and 4 lowest impacts (n=51 offices)		Without offices with 5 highest and 5 lowest impacts (n=49 offices)	
	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)
Program Management										
Emphasis on moving clients into jobs quickly	720	0.000	637	0.000	601	0.000	485	0.003	399	0.011
Emphasis on personalized client attention	428	0.000	385	0.000	354	0.002	305	0.006	267	0.011
Closeness of client monitoring	-197	0.110	-115	0.379	-71	0.568	-32	0.788	-80	0.477
Staff caseload size	-4	0.003	-4	0.011	-3	0.028	-2	0.089	-2	0.147
Frontline staff/supervisor inconsistency about service technology	-159	0.102	-162	0.120	-171	0.096	-191	0.054	-153	0.103
Frontline staff inconsistency about service technology	124	0.141	79	0.348	42	0.596	-24	0.749	-34	0.611
Service Differential										
Basic education	-16	0.017	-16	0.018	-13	0.031	-13	0.025	-9	0.063
Job search assistance	1	0.899	2	0.851	4	0.715	13	0.230	18	0.102
Vocational training	7	0.503	4	0.745	7	0.519	7	0.541	17	0.089
Economic Environment										
Unemployment rate	-94	0.004	-77	0.023	-65	0.052	-57	0.078	-47	0.129
Mean Program Impact on Earnings	879	0.000	866	0.000	847	0.000	845	0.000	818	0.000
Mean Counterfactual	4,871	0.000	4,843	0.000	4,844	0.000	4,852	0.000	4,881	0.000
Impact as Percent of Counterfactual	18.0		17.9		17.5		17.4		16.8	

Note: a: Coefficient values are in 1996 dollars.

Table D3

Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Offices with Early Random Assignment

Program Characteristic	Full Sample (n=59 offices)		Without offices with Early Random Assignment (n=42 offices)	
	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)[b]
Program Management				
Emphasis on moving clients into jobs quickly	720	0.000	455	0.010
Emphasis on personalized client attention	428	0.000	204	0.176
Closeness of client monitoring	-197	0.110	-46	0.775
Staff caseload size	-4	0.003	-6	0.098
Frontline staff/supervisor inconsistency about service technology	-159	0.102	-70	0.601
Frontline staff inconsistency about service technology	124	0.141	155	0.343
Service Differential				
Basic education	-16	0.017	-21	0.027
Job search assistance	1	0.899	-2	0.861
Vocational training	7	0.503	11	0.398
Economic Environment				
Unemployment rate	-94	0.004	-64	0.175
Mean Program Impact on Earnings	879	0.000	1,134	0.000
Mean Counterfactual	4,871	0.000	4,796	0.000
Impact as Percent of Counterfactual	18.0		23.6	

Notes:

a. Coefficient estimates are in 1996 dollars.

b. Unlike values reported in most other tables in this paper, the p-values listed here are not computed from robust standard errors due to the small sample size at Level Two that resulted from the restriction imposed in this sensitivity test.

Table D4

Sensitivity Tests of the Relationships between *Program Characteristics* and Program Impacts: Deleting Offices with the Highest and Lowest Values of Selected Independent Variables

Program Characteristic	Full Sample (n=59 offices)		Without offices with 2 highest and 2 lowest values of the Independent Variable (n=55 offices)		Without offices with 3 highest and 3 lowest values of the Independent Variable (n=53 offices)		Without offices with 4 highest and 4 lowest values of the Independent Variable (n=51 offices)		Without offices with 5 highest and 5 lowest values of the Independent Variable (n=49 offices)	
	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)	Regression Coefficient[a]	Statistical Significance of Coefficient (p-value)
Program Management										
Emphasis on moving clients into jobs quickly	720	0.000	881	0.000	832	0.000	860	0.000	893	0.000
Emphasis on personalized client attention	428	0.000	519	0.001	435	0.005	464	0.004	441	0.008
Staff caseload size	-4.3	0.003	-3.7	0.021	-3.3	0.048	-2.7	0.094	-3.1	0.119
Service Differential										
Basic education	-16	0.017	-10	0.141	-12	0.264	-11	0.391	-14	0.315
Economic Environment										
Unemployment rate	-94	0.004	-71	0.048	-59	0.128	-89	0.019	-116	0.025

Note:

a: Coefficient values are in 1996 dollars.

Table D5
Sensitivity Tests of the Relationships between *Client Characteristics* and Program Impacts: Deleting Riverside GAIN and Portland NEWWS

Client Characteristic	**Full sample (n=59 offices)**		**Without Riverside GAIN (n=55 offices)**		**Without Portland NEWWS (n=52 offices)**		**Without Riverside GAIN *and* Without Portland NEWWS (n=48 offices)**	
At Random Assignment the Sample Member:	**Regression Coefficient**[a]	**Statistical Significance (p-value)**	**Regression Coefficient**[a]	**Statistical Significance (p-value)**	**Regression Coefficient**[a]	**Statistical Significance (p-value)**	**Regression Coefficient**[a]	**Statistical Significance (p-value)**[b]
Was a high school graduate or had a GED	653	0.001	565	0.003	618	0.002	513	0.002
Had one or fewer children (left-out)								
Had two children	301	0.160	250	0.267	388	0.061	348	0.039
Had three or more children	591	0.003	459	0.022	639	0.003	496	0.014
Had a child under six	34	0.841	98	0.573	-51	0.771	-9	0.961
Was less than 25 years old	206	0.557	84	0.814	221	0.541	88	0.816
Was 25 to 34	105	0.707	5	0.988	69	0.811	-31	0.926
Was 35 to 44	305	0.376	180	0.615	219	0.540	84	0.803
Was 45 or older (left-out)								
Was White, non-Hispanic (left-out)								
Was Black, non-Hispanic	-178	0.369	-153	0.452	-91	0.609	-26	0.886
Was Hispanic	-213	0.527	-149	0.679	-250	0.473	-180	0.487
Was Native American	-696	0.115	-506	0.263	-967	0.039	-750	0.233
Was Asian	353	0.560	122	0.817	-520	0.375	-36	0.952
Was some other race/ethnicity	726	0.487	876	0.390	916	0.472	1,127	0.372
Was a welfare applicant	-145	0.532	-219	0.383	-140	0.553	-181	0.453
Had received welfare continuously for the past 12 months	444	0.085	217	0.196	426	0.120	283	0.100
Had zero earnings in the past year (left-out)								
Had earned $1 to $2499	-186	0.222	-193	0.223	-120	0.449	-119	0.539
Had earned $2500 to $7499	72	0.787	104	0.705	143	0.608	192	0.402
Had earned $7500 or more	22	0.965	-330	0.488	70	0.894	-288	0.297
Mean Program Impact on Earnings	879	0.000	700	0.000	765	0.000	565	0.000
Mean Counterfactual	4,871	0.000	4,946	0.000	4,794	0.000	4,861	0.000
Impact as Percent of Counterfactual	18.0		14.2		16.0		11.6	

Notes:

a. Coefficient values are in 1996 dollars.

b. Unlike values reported in most other tables in this paper, the p-values listed here are not computed from robust standard errors due to the small sample size at Level Two that resulted from the restriction imposed in this sensitivity test.

Table D6
Sensitivity Tests of the Relationships between *Client Characteristics* and Program Impacts: Deleting Offices with the Highest and Lowest Impact Estimates

Client Characteristic	**Full Sample (n=59 offices)**		**Without offices with 2 highest and 2 lowest impacts (n=55 offices)**		**Without offices with 3 highest and 3 lowest impacts (n=53 offices)**		**Without offices with 4 highest and 4 lowest impacts (n=51 offices)**		**Without offices with 5 highest and 5 lowest impacts (n=49 offices)**	
At Random Assignment the Sample Member:	**Regression Coefficient[a]**	**Statistical Significance (p-value)**	**Regression Coefficient[a]**	**Statistical Significance (p-value)**	**Regression Coefficient[a]**	**Statistical Significance (p-value)**	**Regression Coefficient[a]**	**Statistical Significance (p-value)**	**Regression Coefficient[a]**	**Statistical Significance (p-value)**
Was a high school graduate or had a GED	653	0.001	625	0.001	619	0.001	620	0.002	591	0.003
Had one or fewer children (left-out)										
Had two children	301	0.160	310	0.151	295	0.180	272	0.223	260	0.259
Had three or more children	591	0.003	638	0.002	589	0.004	537	0.007	487	0.015
Had a child under six	34	0.841	-8	0.965	27	0.875	48	0.780	61	0.732
Was less than 25 years old	206	0.557	299	0.400	301	0.409	261	0.480	229	0.553
Was 25 to 34	105	0.707	161	0.573	211	0.472	221	0.456	213	0.499
Was 35 to 44	305	0.376	361	0.304	389	0.282	382	0.297	348	0.371
Was 45 or older (left-out)										
Was White, non-Hispanic (left-out)										
Was Black, non-Hispanic	-178	0.369	-201	0.324	-188	0.359	-183	0.374	-167	0.425
Was Hispanic	-213	0.527	-237	0.499	-217	0.539	-191	0.589	-159	0.676
Was Native American	-696	0.115	-529	0.224	-458	0.304	-475	0.348	-401	0.429
Was Asian	353	0.560	-447	0.465	-243	0.691	-180	0.771	228	0.660
Was some other race/ethnicity	726	0.487	1,237	0.261	1,470	0.214	1,110	0.342	1,720	0.164
Was a welfare applicant	-145	0.532	-125	0.597	-106	0.659	-134	0.589	-178	0.491
Had received welfare continuously for the past 12 months	444	0.085	480	0.064	449	0.083	467	0.076	390	0.130
Had zero earnings in the past year (left-out)										
Had earned $1 to $2499	-186	0.222	-150	0.329	-123	0.420	-144	0.352	-125	0.431
Had earned $2500 to $7499	72	0.787	135	0.643	180	0.507	172	0.528	202	0.467
Had earned $7500 or more	22	0.965	14	0.978	-38	0.938	-54	0.913	-138	0.778
Mean Program Impact on Earnings	879	0.000	866	0.000	847	0.000	845	0.000	818	0.000
Mean Counterfactual	4,871	0.000	4,843	0.000	4,844	0.000	4,852	0.000	4,881	0.000
Impact as Percent of Counterfactual	18.0		17.9		17.5		17.4		16.8	

Note: a. Coefficient values are in 1996 dollars.

Table D7

Sensitivity Tests of the Relationships between *Client Characteristics* and Program Impacts: Deleting Offices with Early Random Assignment

Client Characteristic	Full sample (n=59 offices)		Without Offices with Early Random Assignment (n=42 offices)	
At Random Assignment the Sample Member:	**Regression Coefficient[a]**	**Statistical Significance (p-value)**	**Regression Coefficient[a]**	**Statistical Significance (p-value)[b]**
Was a high school graduate or had a GED	653	0.001	913	0.000
Had one or fewer children (left-out)				
Had two children	301	0.160	153	0.473
Had three or more children	591	0.003	457	0.053
Had a child under six	34	0.841	52	0.812
Was less than 25 years old	206	0.557	-320	0.463
Was 25 to 34	105	0.707	-177	0.631
Was 35 to 44	305	0.376	-103	0.784
Was 45 or older (left-out)				
Was White, non-Hispanic (left-out)				
Was Black, non-Hispanic	-178	0.369	-367	0.117
Was Hispanic	-213	0.527	-235	0.404
Was Native American	-696	0.115	-612	0.486
Was Asian	353	0.560	-707	0.260
Was some other race/ethnicity	726	0.487	845	0.494
Was a welfare applicant	-145	0.532	28	0.948
Had received welfare continuously for the past 12 months	444	0.085	625	0.002
Had zero earnings in the past year (left-out)				
Had earned $1 to $2499	-186	0.222	-300	0.200
Had earned $2500 to $7499	72	0.787	52	0.861
Had earned $7500 or more	22	0.965	654	0.070
Mean Program Impact on Earnings	879	0.000	1,134	0.000
Mean Counterfactual	4,871	0.000	4,796	0.000
Impact as Percent of Counterfactual	18.0		23.6	

Notes:

a. Coefficient values are in 1996 dollars.

b. Unlike values reported in most other tables in this paper, the p-values listed here are not computed from robust standard errors due to the small sample size at Level Two that resulted from the restriction imposed in this sensitivity test.